

**INTERNATIONAL EVIDENCE, NATIONAL POLICY
AND EDUCATIONAL PRACTICE**
Making better use of international comparisons in education

Robin Alexander

Universities of Cambridge and York
Chair, Cambridge Primary Review Trust

Introduction¹

I'd like to talk about a belief that surfaces with increasing regularity in the educational policy discourse of Britain and indeed many other countries. It is the belief that we will solve the nation's educational problems not by investigating their distinctive pathology and revealing their cultural, historical and political determinants, thus arriving at explanations and solutions that have a chance of being valid, honest and appropriate, but by copying or adapting the policies of those countries whose students outperform our own on a few narrow and contested measures of educational achievement. In this process of policy grafting or transplant it appears not to matter whether the countries in question are comparable to our own or utterly unlike it, or whether they are passable democracies in which educational debate is encouraged, or autocratic regimes where debate and dissent are suppressed.

Or rather, the currently fashionable unit of comparison is not 'countries', still less 'cultures', but 'jurisdictions'; for 'jurisdictions' allows governments and their advisers to smuggle barely sustainable comparisons between a country like Britain with a population of 60 million and 25,000 schools, a city state like Singapore with just 5 million inhabitants and 350 schools and a Californian charter school system with just 40 schools. Thus, in the letter of 11th June with which England's Secretary of State for Education launched his proposals for a new national curriculum, the phrase 'high performing jurisdictions' or something like it is used three times in the first four paragraphs and many more times in the rest of the letter by way of justification for his claims. Many questions are begged by that phrase 'high performing jurisdictions', by the evidence that supposedly informs it, and by the policies that it is used to justify. In this keynote I shall invite you to think about some of them.

Why do we compare?

In many countries, especially those that are members of OECD, the business of comparing educational systems and outcomes has become a political and media obsession, generating celebration in some quarters and panic and blame in others. In between the extremes of panic and celebration are those countries – and sadly there aren't many of them – that ponder international evidence quietly, carefully and critically before deciding what action, if any, to take.

¹ This paper was first given, under the title *International evidence, national policy and classroom practice: questions of judgement, vision and trust*, as the closing keynote presentation at the Third Van Leer International Conference on Education in Jerusalem in May 2012. In modified form and with a different title it became a keynote at the July 2012 international teacher education conference at Canterbury Christ Church University.

This preoccupation with the educational performance of other countries might seem novel but it isn't, for international educational comparison and the globalisation of educational ideas are as old as education itself. Thus you will find ...

... Pestalozzi mingling with Tagore, Krishnamurti and the Elmhursts in both English and Indian progressive education, and John Dewey's ideas turning up briefly in China, the Soviet Union and Turkey as well more lastingly in England and the United States. Meanwhile, both the German *Gymnasium* and the American high school help shape the development of Russian schooling. British officials take the *Ecole Normale* from France to England and India. Jan Komensky (Comenius) journeys tirelessly from Moravia to Prague, Heidelberg, Berlin, Paris, Stockholm, London, Amsterdam and points between and beyond, and corresponds with the founding fathers of Massachusetts, and his principles of common vernacular schooling and carefully graduated whole class teaching, not to mention his textbooks, embed themselves deeply and lastingly in the pedagogy of many countries of central, eastern and northern Europe. And during the nineteenth century the monitorial systems of Bell and Lancaster seed themselves just about everywhere from their probable roots in what was then Madras. (Alexander 2001, 171-2).

If international comparison and what some call 'policy borrowing' have been around a long time, so too have warnings about their misuse. Here is British educationalist, administrator and university vice-chancellor Michael Sadler, writing over a century ago:

In studying foreign systems of education we should not forget that the things outside the schools matter even more than the things inside the schools, and govern and interpret the things inside ... No other nation, by imitating a little bit of German organisation, can thus hope to achieve a true reproduction of the spirit of German institutions ... All good and true education is an expression of national life and character ... The practical value of studying in a right spirit and with scholarly accuracy the working of foreign systems of education is that it will result in our being better fitted to study and understand our own. (Sadler 1900, 50)

So what would Sadler make of the definition of an 'ideal school' by OECD's Andreas Schleicher (2009)? Such a school, said Schleicher, would include 'little bits' of Finland, Japan, England, Israel, Norway, Canada, Belgium and Germany in the way it develops students' individual strengths, gets teachers to co-operate, sets clear performance targets, celebrates discourse and helps students to learn from their mistakes, unites kindergarten and primary education, focuses local support where it is needed, values the arts, and devolves responsibilities to students. (Here's a test for you: which 'little bit' of Schleicher's ideal school comes from England?)

Sadler's objection to international 'cherry-picking' is principled but it fails to take account of the reality of the global educational traffic that I referred to a moment ago. Yet his basic contention is surely right. National education systems are embedded in national culture, and indeed education is one of the main ways that a culture is mediated, transmitted and sustained; so 'no educational policy or classroom practice can be properly understood except by reference to the web of inherited ideas and values, habits and customs, institutions and world views, that make one country distinct from another.' (Alexander 2001, 5).

This nexus of education, culture and history means that while, as Sadler says, 'All good and true education is an expression of national life and character', there's also a downside, for as American educationist and administrator Ernest Boyer noted, 'A report card on public

education is a report card on the nation. Schools can rise no higher than the communities that support them.’ (Boyer 1983, 6) Or, as Basil Bernstein more succinctly put it 40 years ago: ‘Education cannot compensate for society.’ (Bernstein 1970, 344).

So if a national education system is not – as Sadler puts it - ‘good and true’, then we must look to society as well as our schools, not merely copy the schooling of others. Thus it is that the gap in student achievement between high and low-attainers is wider in Britain and the United States than many other OECD countries in part because it coincides with disparities in income, health and wellbeing which are also wider in those two countries than elsewhere. As Wilkinson and Pickett (2010) demonstrate in their groundbreaking international epidemiological study *The Spirit Level*, unequal societies have unequal education systems and unequal educational outcomes. That’s a finding – or a maxim - worth pondering.

Obvious all this may be, yet it’s not obvious enough to those who find it more expedient to copy the policies of other countries, systems or ‘jurisdictions’ than to examine their own countries’ condition and needs. The word ‘jurisdictions’ tacitly sanctions this neglect, for it strips a country of the complexities of culture, values, social structure, politics and demography and reduces it to tidy legalities. But these are the very complexities with which we must engage if we are to understand education elsewhere, let alone explain why one country outperforms others.

You may be surprised that I’ve got this far without mentioning PISA or TIMSS, for to many people these days PISA and TIMSS are the only international evidence that matters, and international comparison means enviously comparing ourselves with PISA top performers Hong Kong, Singapore, Japan, Korea or Shanghai-China, or following the PISA tourist trail to Finland. Whether the test performance of 15-year olds in literacy, mathematics and science is a sufficiently robust indicator of what Sadler called ‘good and true’ education is rarely considered in this age of PISA panic and PISA tourism, but it’s a question that needs considering, not least when we read ludicrous newspaper headlines like this:

TRAVESTY OF OUR ‘STAGNATING’ SCHOOLS: OECD CONDEMNS BRITISH EDUCATION WHICH IS NOW INFERIOR TO ESTONIA’S ... (Loveys 2010)

or this:

DISASTROUS RESULTS FOR WALES IN PISA TEST: ‘It is simply not acceptable for Estonia to be outperforming Wales ...’ said Liberal Democrat spokeswoman ... (News Wales 2010)

Note the imperialist and possibly racist assumption that Estonia’s proper place in the world order is well below that of England and Wales. And if you think such hysterical headlines are unique to the UK, consider these:

PISA TEST RESULTS SHOW COLOSSAL FAILURE OF ISRAEL SCHOOLS (Hartman 2010).

PISA DESASTER: WAS MACHT DIE SACHSEN BESSER? (Klostermann and Wolff 2010)

LA FRANCE PARALYSÉE DEVANT SES MAUVAIS RÉSULTATS SCOLAIRES (Rollot and de Vergès 2007)

There in a nutshell we have it: the performance of a group of 15 year olds, at just one brief moment during their long educational journey, in selected aspects of just three school subjects, albeit extremely important subjects, demonstrates the failure of an entire national education system. What a counsel of despair. What a ludicrous level of inference far, far beyond what the data allow. What a denial of the true potential of international comparison. And what a heavy burden of responsibility to place on those 15 year olds and their teachers - especially when they see the other headline:

SHANGHAI KIDS THE WORLD'S SMARTEST (Bloom 2010).

International comparisons: questions of evidence

The way the discourse of international comparison is dominated by the international achievement surveys and the accompanying media and political hysteria requires us to think more deeply about evidence. In comparing ourselves with others have we got the balance of evidence right? Are we taking too much notice of some kinds of evidence and too little of others?

Among the many typologies of comparative international research one of the more useful in this context has come from the National Research Council (NRC) of the US National Academies, which for many years has advised and commented on US participation in the international student achievement surveys. In 2003, NRC's Board on International Comparative Studies in Education (BICSE) identified three main types of comparative study (NRC 2003).

Type I includes the large-scale international student achievement studies such as TIMSS and PISA, not to mention FIMS, SIMS, FISS, SISS, TIMSS-R, PIRLS, ICCS, SITES, TEDS-M and - oh yes - WOEDAM. That's OECD's latest test, to be administered before you leave this room today: it stands for What On Earth Do these Acronyms Mean? These surveys are typically quantitative and are referenced by a limited number of indicators and measures of learning outcome and, latterly, of context and input.

The usual *Type II* paradigm is the desk-based extrapolation from existing international data in order to propose policy options in specific areas like school leadership, teacher training, curriculum or assessment. The best known examples are the three McKinsey reports (Barber and Mourshed 2007, Mourshed, Chijioke and Barber 2010, Barber, Whelan and Clark 2011).

Type III studies include the majority of work in the published corpus of academic comparative education. They range from descriptive accounts of individual education systems to close-grained comparative studies of school and classroom life and the cultural and historical forces that shape it, uncovering similarities as well as differences. Examples of Type III comparative research are Joseph Tobin's ethnographic studies of pre-school education in China, Japan and the United States (Tobin *et al* 1989, 2009), Marilyn Osborn's and Patricia Broadfoot's exploration of the experience of being a learner in England, France and Denmark (Osborn *et al* 2003) and Jin Li's re-examination of 'Chinese' and 'western' models of learning (Li 2012). (Modesty forbids mention of my own work). The scale, range, methods and quality of Type III studies are variable, but their common feature is this: many of them have significant policy applications, as the NRC report notes, but the imperatives of policy are not their starting point. Rather, their goal is the advancement of educational understanding for its own sake. Following Michael Sadler, they compare to learn, not to copy.

But here's the crunch. The 2003 NRC report notes that while the majority of published comparative education studies are Type III, it is the Type I and II studies that receive most of the funding, political patronage and publicity, and the funding difference can be truly vast. In stark contrast, Type III studies scrape together what they can from charities and research councils and are disseminated through the decidedly low-profile educational press. Yet, the NRC report goes on:

Although they vastly outnumber Type I and Type II studies, Type III studies do not usually come to the attention of policy makers or the public. This is a loss, since many are rich in narrative detail and paint a more engaging and provocative portrait of education in other countries than do the summary bar charts and graphs typical of the larger studies. Ethnographic and case studies, in particular, can explore cultural context in depth and ... elucidate the way education is organised and understood in different cultures. (NRC 2003, 23-4).

It hasn't always been like this. The linguistic anthropologist Shirley Brice Heath reminds us that in the United States and Britain, and no doubt in other countries too, there have been times when ethnographers and anthropologists have been eagerly enlisted and generously supported to provide evidence to inform the development of policy, for example during periods of rapid social change, national restructuring or international conflict (Heath and Street 2008). For the moment, however, governments prefer to look elsewhere.

From international evidence to national policy: examples from England

What the NRC report is saying, if we can express the matter even more bluntly, is that in pursuit of what they call 'evidence-based policy', governments choose to ignore the larger part of the international evidence that is available to them, including evidence that could give them the insights, explanations and policy options they need. Let me give you examples of this evidential selectivity at work from England – not because England is especially at fault but simply because I know it best. International delegates at this conference can judge how far the English experience resonates with their own.

In 1996 Ofsted, England's national school inspectorate, commissioned and published *Worlds Apart?* This report reviewed England's performance in the international student achievement surveys up to the mid 1990s (FIMS, SIMS, FISS, SISS and IAEPS 1, 2 and 3), proposed reasons for England's generally poor performance, and identified solutions extrapolated from the top performing systems in central Europe and Asia. It recommended 'high quantities of whole-class interactive instruction ... [and] the use of the same textbooks by all children,' (Reynolds and Farrell 2006, 55). The report's thesis, deploying the uncomplicated attributions of cause and effect that some politicians prefer, was that whole class teaching raises standards in literacy and numeracy which in turn boosts national economic performance. If only it were that simple (Robinson 1999, Wolf 2002).

In 2010, the government-endorsed report *Could do Better?* resumed the narrative, using PISA 2009 as its baseline. This time, the key to raising standards was the prescribed curriculum because, the report asserted, 'in all PISA high-performing systems the fundamentals of subjects are strongly emphasised, have substantial time allocation and are the focus of considerable attention.' (Oates 2010). The report then listed 13 'control factors' - national tests, state-approved textbooks, school inspections, closely-aligned teacher training and so on – through which the report claimed that governments in PISA high-performing systems ensure that teachers toe the line – a claim which some of the governments in question might

wish to challenge. Even as I speak, this report is informing the UK government's decisions about national curriculum reform in England (DfE 2010, 2011a).

Third, there's the phenomenon of McKinsey. In 2007, this multi-billion dollar international management consultancy published the first of three reports from a team headed by Michael Barber, formerly of Prime Minister Blair's back office, on what those education systems that topped the PISA 2003 league table could teach the rest of us. Its provocative title was *How the World's Best-Performing Education Systems Come Out on Top*. From its analysis of the world's top ten PISA systems McKinsey concluded: 'Three things matter most: (1) getting the right people to become teachers, (2) developing them into effective instructors, and (3) ensuring that the system is able to deliver the best possible instruction for every child.' (Barber and Mourshed 2007, 2).

Did we really need McKinsey to tell us that? In the course of its expensive quest for this statement of the obvious the same report delivered further stunning insights such as: 'The quality of an education system cannot exceed the quality of its teachers', 'The only way to improve outcomes is to improve instruction' and 'High performance requires every child to succeed.' To such multiple tautologies McKinsey added linguistic gems like: 'Top-performing systems leverage a substantial and growing knowledge about what constitutes effective school leadership to develop their principals into drivers of instruction.' (Barber and Mourshed 2007, 4 and 30).

Thus we have three reports, all enjoying classic Type II political patronage (in McKinsey's case by many more governments than that of the UK), all starting from similar Type I datasets yet each extracting different recipes for school improvement: more whole class teaching based on standardised textbooks (Reynolds); better teaching, teacher training and school leadership (McKinsey); a drastically pared-down curriculum which concentrates on what is deemed 'essential knowledge in the key subject disciplines', backed by 'control factors' to prevent teachers from making up their own minds about what knowledge is essential. (Oates as cited in the remit for England's 2010-12 national curriculum review, DfE 2010).

From international evidence to national policy: questions of judgement and trust

What have England's policymakers done with these Type II international studies? The 1996 *Worlds Apart* report's advocacy of whole class teaching and standardised textbooks helped shape the national literacy and numeracy strategies introduced by the Labour government in 1998-9 which for a decade were imposed on every primary school, primary teacher and primary student in England. The 2010 *Could do Better* report was used to justify the current government's decision to slim down England's national curriculum to 'essential knowledge' in English, mathematics and science and perhaps a few other subjects, backed by phonics checks and government-approved phonics materials to ensure that all primary teachers follow the official line on literacy teaching. And where would the government seek guidance on what constitutes 'essential knowledge'? On this the 11 June statement of Secretary of State Michael Gove is startlingly unambiguous, asserting (my italics) that we must 'ensure that our children master the essential core knowledge which *other nations* pass on to their pupils' (DfE 2012, 2).

The impact of McKinsey in England has been more diffuse. It can be seen in the UK government's raising of the academic entry bar for trainee teachers, though nowhere near as high as Finland, the country which is usually cited as the model for this particular reform. However, while saying that good teaching is what makes the difference, McKinsey offers

nothing of substance on pedagogy so its reports have no *classroom* purchase. Instead they pin their faith in the levers and controls of systemic reform, in more and more legislation and regulation: not surprisingly, for Michael Barber was not only political adviser to Prime Minister Blair, but was also the architect of those national literacy and numeracy strategies, the massively expensive intervention through which Blair's government sought to raise standards in England's primary schools, to questionable effect. So Barber had a point to prove. (The evidence on the impact of this particular exercise in major systemic reform is examined in the final report of the Cambridge Primary Review (Alexander 2010b, 471-4).

But the more fundamental problem is that when Type II comparative studies are linked to the imperatives of policy they can be highly selective in their use of evidence. For example, the prescribed curricula of PISA high performers certainly emphasise essential knowledge as the 2010 Oates report claims – what prescribed curriculum doesn't? – but they do *not* advocate a narrow focus on only those subjects that PISA tests. On the contrary, the UK government has itself published evidence to show that the prescribed curricula in high performing systems require breadth both across the curriculum as a whole and within each of its main subjects (DfE 2011). And in the UK there's even a consistent line of school inspection data, of which the government has been aware for two decades, that the schools whose students do best in the national tests of literacy and numeracy at age 11 don't concentrate on literacy and numeracy to the exclusion of all else but embed them in a curriculum which is broad, balanced and well-managed. Similarly, from the United States last year we had a report showing that the arts not only motivate and engage children but also help raise measured standards in literacy and numeracy (President's Committee on the Arts and Humanities 2011). The report's honorary chair was First Lady Michelle Obama. Did her husband pause in his 'race to the top' to talk to her about it?

Learning in one area enhances learning in others. It's an obvious and powerful truth. So why in political circles doesn't it cut any ice? Well, in the context of policy such a truth is inconvenient. The instinctive and preferred policy response to falling standards in literacy and numeracy is to hammer literacy and numeracy to the exclusion of all else. That, governments presume, is the kind of tough and uncompromising response that the electorate wants to hear. Perhaps, in remaining pointedly silent on the research evidence about the connections between music and literacy, or between classroom argumentation and scientific reasoning, governments also underestimate the electorate's intelligence.

As to the Type I datasets which provide Type II studies with their benchmarks, from the methodologically shaky IEA and IAEP studies of the 1970s to the much more sophisticated 2009 PISA, these have become more reliable and culture-fair, and the most recent PISA commentaries produced by OECD demonstrate considerable understanding of the challenges of cross-cultural testing and the inferential and explanatory caveats that need to be entered and heeded. (OECD 2010a, 2010b, 2010c, 2010d, 2011a, 2011b).

In this sense, the problem isn't PISA itself but what people do with the data that PISA provides. Hence my particular concern about the Type II data extrapolations and the disproportionate influence they exert. For the political attraction of Type II studies is that they select, mediate, repackage and re-interpret the research of others, presenting it in a form that they believe policymakers will find palatable. But being gatekeepers rather than creators of evidence, authors of Type II studies are acutely vulnerable to the charge of methodological and even political bias. The McKinsey reports, in particular, draw on an astonishingly narrow and partisan range of sources. A McKinsey bibliography reads like the membership list of a very exclusive club.

Since, in spite of such failings McKinsey's three education reports are proving so influential internationally, we might press the critique further. David Raffe (2011) argues that the second McKinsey report remains locked within the policy 'bubble'; that it is a particularly flawed example of policy borrowing; that it ignores educational processes and contexts; and that it applies exceptionally narrow criteria of systemic educational success. But the most devastating criticisms come in a recent article by Frank Coffield (2012). His central charge is that McKinsey's analysis is culpably mono-factorial, failing to give due weight to the socio-economic factors that have a significant impact on students' motivation and attainment, especially in a country whose economic and social disparities are as great as they are in Britain. The first McKinsey report, continues Coffield, is methodologically flawed, 'disablingly selective' in its data and explanatory frame, superficial in its account of 'best practice' and how this can be disseminated, and seduced by its own rhetoric on leadership. Coffield likes the second McKinsey report better than the first, though still not a lot, but castigates it for an impoverished view of teaching and learning, a thin evidence base, implausible arguments about the mechanisms and processes of school improvement, technocratic and authoritarian language and a pervasive neglect of culture and political context. Coffield sees McKinsey as

the work of 'global' policy analysts, remote from the complexities of classrooms and the discomfiting findings of researchers which pose such difficulties for politicians in search of quick 'transformations' of school systems before the next election. They espouse a ... model of schooling ... characterised by relentless pressure, competition, line managers, customer services, data for performance management, accountability and value for money; and professional autonomy for teachers only when granted by the centre ... Their notion of teaching is narrowly conceived and technocratic ... Their model remains unsophisticated, impracticable and undemocratic ... Their recommendations are educationally and socially dysfunctional and should not be part of school reform in a democracy. (Coffield 2012).

Strong stuff indeed, but in my view not undeserved. Stepping back from reports like those from McKinsey, Ofsted and Oates, there are wider concerns about the current official discourse of international comparison.

First, it has become dangerously over-politicised. Between 1997 and 2010 TIMSS and PISA data were enlisted to prove the success of the UK Labour government's drive to raise standards in English schools. Then in 2010 there was a general election which Labour lost. Their successors immediately used PISA 2009 to show that far from rising, student performance had 'plummeted' (Young, 2010). In Parliament, the Education Secretary summarised England's PISA performance in nine words: 'Literacy, down; numeracy, down; science, down: fail, fail, fail'. This pessimistic reading of PISA was then used to justify major policy changes of which cutting back the national curriculum is just one.

However, after re-analysing the TIMSS and PISA data John Jerrim (2011) of London University's Institute of Education concluded that they neither justify such alarmist claims nor provide a safe evidential basis for sweeping policy changes. Earlier studies commissioned by the Cambridge Primary Review had come to a comparable conclusion about Labour's overly optimistic spin on the international and national test data (Tymms *et al* 2010, Whetton *et al* 2010).

It seems only fair to ask how far, and in how many countries, PISA is being exploited or manipulated for political purposes.

Second, PISA assesses the attainment of 15 year olds in aspects of reading, mathematics and science. Its spectrum of 'key competencies', though undoubtedly essential, is also quite limited. OECD itself acknowledges this, saying that PISA covers just 'some of the knowledge and skills that are essential for full participation in society'. (OECD 2012). OECD is right, and governments should pay attention: PISA tests what it tests. PISA is not a proxy for the whole of a child's education. It is not a sufficient basis for describing an education system *as a whole* as 'high performing' or 'failing'. But McKinsey and other influential Type II studies are premised on precisely this inference, as of course are newspaper headlines of the kind I quoted earlier.

Third, those who claim a simple cause-effect relationship between, on the one hand, whole class teaching, standardised textbooks or a curriculum prescribing 'essential knowledge in key subjects' and, on the other hand, the capacity to outperform other countries in TIMSS and PISA, risk false correlation, or the philosophers' 'fallacy of division'. X may well be a common feature of high-performing education systems a, b, c, d and e, but that doesn't demonstrate a cause-effect relationship between feature and performance. And if x is also a common feature of *low*-performing systems g, h, i, j and k, then the claimed correlation is clearly inadmissible. In fact, to stay with these examples, whole class teaching, standardised textbooks and a prescribed curriculum concentrating on what is deemed 'essential knowledge in the key subjects' are all international defaults, featuring in low-performing systems as well as high (Benavot *et al* 1991, Benavot 2008, Alexander 2008, chapter 2).

Fourth, although much is made of Finland, politically inconvenient truths about Finland's success may be ignored. Scandinavia's own experts tell us that Finland's TIMSS and PISA performance reflects a culture that has an exceptionally high regard for literacy, a highly trained, well trusted and autonomous teaching profession, an unshakeable commitment to social and educational equity, a successful comprehensive school system, and close alignment not of so-called 'control factors' like testing, textbooks and inspection but of public policy in education, the economy, employment and social welfare. (Lyytinen 2002, Fredriksson 2006, Sahlberg 2011). And there's something else: Pasi Sahlberg's argues that the political and professional climate must above all be characterised by *trust*, something which is in very short supply indeed in British political life. Further, trust must be omni-directional: governments must trust teachers, and we must all be able to trust our governments and the evidence they cite in support of their policies. 'Trust can only flourish in an environment that is built upon honesty, confidence, professionalism and good governance ... Trusting schools and teachers is a consequence of a well-functioning civil society.' (Sahlberg 2011, 130-1, my italics). In this matter, Britain's complacent and imperfect democracy has a very long way to go.

Fifth, and perversely, having praised the Scandinavian model some governments then go on to copy the Anglo-Saxon one, hence high stakes testing, punitive inspection and the marketisation of schooling – strategies which Finland has explicitly rejected and which, as has been shown by Diane Ravitch, Sharon Nichols, David Berliner and others in the United States (Ravitch 2010, Nichols and Berliner 2007, Ravitch 2012) and by the Cambridge Primary Review in England (Alexander 2010b, Alexander *et al* 2010), generate considerable collateral damage while not necessarily delivering on standards. Why? Is it that political imperatives outweigh all others and policymakers believe they must not only act but be seen to act, and while the evidence suggests the need for more grass-roots support and professional trust, what plays better with voters is the promise of tough action in the form of top-down regulation? Or is it simply that in a contest between evidence and political ideology, ideology always wins? So much for evidence-based policy.

Finally, once we move beyond the restricted range of system and school variables deployed in Type II studies, we encounter wider social, cultural, demographic and economic conditions which directly and massively influence the educational performance of a country's students. While McKinsey rightly concludes that teachers and teaching make a considerable difference, extra-educational factors like country size, per capita GDP, demographic homogeneity and relative equality correlate no less convincingly with PISA performance. A glance at the top end of the PISA league tables shows that the systems in question are mostly small, rich or preferably both. To these factors, according to Wilkinson and Pickett, we must add equality and equity:

Greater equality, as well as improving the wellbeing of the whole population, is also the key to national standards of achievement and how countries perform in many different fields ... If ... a country wants higher average levels of educational achievement among its school children, it must address the underlying inequality which creates a steeper social gradient in educational achievement. (Wilkinson and Pickett 2010, 29-30)

Interestingly, after PISA 2009 OECD confirmed the significance of equity as a factor in national performance, (OECD 2010c, OECD 2010d, 9-11) though on this vital matter the Type 2 studies I have exemplified are all strangely silent. Perhaps this is because for politicians in Britain and the United States, which are among the most unequal of all the OECD nations and where the gap between rich and poor continues to widen, this is the most inconvenient truth of all. (Alexander 2010a).

Making better use of international comparison

Once one entertains a wider range of explanatory factors in international student achievement the possibilities for genuinely useful international comparison increase dramatically.

For example, in 2007 UNICEF published a study placing the UK at the bottom of a league table of children's wellbeing in 21 rich nations, using indicators for income, health and safety, family and peers, behaviour and risk, and of course education, plus one for the subjective judgement of children themselves. (UNICEF 2007). A follow-up study explored the reasons for Britain's poor showing in the 2007 study by comparing the UK, Spain and Sweden. (UNICEF 2011). It concluded that a materialistic and commercial culture is much more deeply embedded in the UK and in its parenting practices than in Spain or Sweden, and that this culture reinforces inequality, adversely affects family time and relationships, and overall has a negative impact on children's wellbeing. Since we know that the maps of student attainment and social equity broadly coincide, we have in such studies clues to how schooling, through the values and relationships it fosters as well as through its quality of teaching, might raise the standards of both attainment and wellbeing.

Staying with wellbeing, here's another example. My Cambridge colleague John Gray has recently completed a meta-analysis of international data on adolescent wellbeing. This shows that while Netherlands and Finland both rank high on PISA, Dutch students are happier at school than their Finnish contemporaries (both countries are represented here tonight so I hope this doesn't cause a diplomatic incident). Gray also argues that McKinsey-style comparisons between the UK and Hong Kong or Singapore are pointless because the cultures are so different and the educational systems are not remotely comparable in scale (England has 23,000 schools while Singapore has just 350). He therefore suggests that if England really wishes to use international comparison for policy transfer it should look not

at Hong Kong, Singapore or even Finland, but at the Netherlands, because this country is successful in the interlocking areas of attainment and wellbeing and it is not too different demographically and culturally from the UK. (Gray *et al* 2011).

Another Cambridge colleague takes a different but no less intriguing view. Like myself and John Gray, David Hargreaves sees culture as central to the understanding of differences in educational structure, content, process and outcome, but doesn't follow Gray's line on how we should act on that understanding. For Hargreaves, the world's educational and economic centre of gravity in the 21st century will be China, and since Shanghai-China outperforms other countries in PISA it is essential that we investigate what it is about Chinese culture that accounts for this. This inevitably leads him to the resilience and pervasiveness of Confucianism, whose values he suggests might profitably inform our own schools' collective identity and their approach to self-improvement. (Hargreaves 2011).

Then there's the important line of research currently being pursued by Yariv Feniger and Adam Lefstein (2012) in Israel. Feniger and Lefstein are deeply critical of what they call the 'education systems hypothesis' that informs McKinsey and some other influential studies I have referred to - and indeed the commentaries of PISA boss Andreas Schleicher himself (Feniger and Lefstein 2012, 1, Schleicher and Stewart 2008) - on the grounds that it attributes too much to policy 'levers' like high stakes testing and too little to history and culture. They argue that if randomised control experiments were possible in this area then they would support the alternative 'cultural-historical' hypothesis about what has the greatest impact on student attainment. They test the hypothesis by taking immigrant children - those who were either born in the host country or immigrated there before starting school - and comparing their test scores with those of their indigenous peers from both their host and their source countries. Looking at children of Turkish and Chinese origin, that is children whose source countries perform very differently in PISA, Feniger and Lefstein show that 'the mean achievement of immigrants is closer to students in their country of origin than to students in their host country.' The decisive factor, then, is culture, especially as it bears on parental attitudes, values and expectations, but culture also perhaps as magnified by the pedagogy of the host countries, whose teachers have heard a great deal about high-achieving East Asian students and raise their expectations accordingly.

And so to pedagogy, the missing ingredient in the McKinsey reports and the point at which culture and policy together exert their greatest leverage on student learning processes and outcomes. Over the past decade I have been trying to apply the findings about pedagogy that came from my own international comparative studies in England, France, India, Russia and the United States. These studies culminated in a comparative analysis of classroom talk: a double culmination, because it is through language, and especially spoken language, that both culture and pedagogy are mediated. Out of this work, and linking to the now extensive international body of research on effective classroom talk, I have developed a particular approach to dialogic teaching (Alexander 2008a, 2008b) on which I have worked extensively with teachers in the UK and other countries. Time doesn't permit me to go into detail, but in the context of seeking alternative ways of using international evidence to improve the quality of education the simple maxim I would propose to you is this: *look for the principles that underpin effective educational practices observed elsewhere, don't merely copy the practice*. For once they have been made explicit these principles can then be discussed, debated, domesticated and re-applied - or not. In education, policies and practices are local and context-bound, but their underlying principles - whether of effective teaching, school leadership or good governance - have a chance of being universal.

Which is what some of us explored in Pittsburgh last September, at an international

conference sponsored by the American Educational Research Association (AERA), and organised by Lauren Resnick of Pittsburgh University and Christa Asterhan of the Hebrew University in Jerusalem. This event brought together international researchers on linguistic and cognitive development, pedagogy and classroom talk. The conference judged that we now have a critical mass of evidence demonstrating that classroom talk that is well-structured, reciprocal and cognitively challenging, and which immerses the student in a subject's distinctive conceptual and linguistic architecture and modes of enquiry, reasoning and argument, has a clear and positive impact on student attainment in language, mathematics and science as assessed by conventional tests. (Resnick, Asterhan, Clarke and Hofkens 2012). Further, and critical in relation to my concerns, although the circumstances and practices of schools and classrooms may vary considerably both within and across cultures, the *principles* of educationally productive talk appear to travel without too much difficulty from one classroom and culture to another.

And in England where has this led? Well, I can tell you that the international evidence on spoken language has been formally summarised for the UK government. I even persuaded DfE to hold a high-level seminar on the matter, five months ago, attended by the Minister with videolinks to top researchers in the United States. The case was powerfully made - and ignored. In the national curriculum proposals published on 11 June the profile of spoken English is even lower than in the current national curriculum. Why? The Minister said he didn't want to do anything to encourage 'idle chatter' in class.

Conclusion: where there is no vision ...

These are just a few examples of alternative kinds of learning from international educational comparison. There are many others (Alexander *et al* 1999, 2000). It seems to me that PISA panic, PISA-driven regulation, and the narrowing of the evidential and explanatory frameworks for international comparison all demonstrate not only a catastrophic loss of judgement and trust but also a serious decline in national self-belief. They also reduce our options for learning from international comparison. The trend, therefore, impoverishes our educational vision. And 'Where there is no vision ...' (you know how that proverb ends).²

This is the year of PISA 2012. Here in conclusion are a few questions about vision that we might ask ourselves, and our governments, when once again PISA panic threatens to overwhelm us.

Certain countries consistently do well in the international tests of student attainment. But what else do their students learn? Do these countries provide their children with an education which is about significantly more than passing tests in two or three subjects? And if the wider curriculum in top-performing TIMSS or PISA systems were to be measured as assiduously as reading, mathematics and science are measured, would the same countries still head the league table?

Is it not at least possible that this whole current PISA-led obsession does a grave injustice to those countries, schools and teachers that care no less passionately about educational standards but for whom standards mean more than test performance? Ultimately, therefore, are we doing an injustice, in the otherwise commendable cause of raising standards, to the cause of education itself?

² Proverbs, 29,18. The translators of the King James Bible published in England in 1611 gave us 'Where there is no vision, the people perish.' This was what I refer to and it is apt enough for my purpose. However, I have been reliably informed that the original Hebrew is more correctly translated as 'Where there is no vision, the people become unruly'. High-handed education ministers should take note.

'World class' education is the claim of governments, businesses, schools and universities everywhere, even though by now they must surely know that the phrase has become almost meaningless (Alexander 2010a). Almost meaningless but not quite, for while 'world class' has little substance it signals a clear enough goal: outperforming other schools and countries. The goal is uncomplicatedly supremacist: how, in McKinsey's words, to 'come out on top.' But in a world facing the crises of climate change, resource depletion, over-population, environmental degradation and economic and geopolitical instability, is this really how 'world class' education should be defined? Should we not consider the merits of 'world class' as *sustaining* the world rather than beating it? As fostering international interdependence and co-operation rather than national supremacy? Whatever happened to education's *moral* purposes?

A world class educational vision needs to be matched by a world class approach to defining and assessing educational performance, whether of students, schools or nations. Is our account of educational performance consistent with what education should be about? I don't think so. It is no longer acceptable to attend only to what can be measured, or to presume that what is measured can validly serve as proxies for education as a whole, or to claim that our children deserve a broad and rich education but then give the lie to this claim by assessing only what is testable in literacy, mathematics and science and ignoring the rest.

In this context note the assertion in the first McKinsey report: 'Top-performing systems recognise that they cannot improve what they do not measure' (Barber and Mourshed 2007, 36). Really? We cannot improve what we do not measure? Try telling that not just to any teacher or parent but also to any artist, writer, philosopher, religious leader or ordinary citizen who strives to improve their thinking, their understanding, their conduct, their lives, and the lives of others - or for that matter anyone who lives outside the bubbles of policy and performativity. Did we learn nothing from the fierce debates about instructional and expressive educational objectives during the 1960s and 1970s? (Popham *et al* 1969, Eisner 1997). Why, when we come to define and assess educational performance, are so many of us still stuck in the nineteenth century?

I stress one more time that the problem isn't PISA or TIMSS but what people do with them; though perhaps, in view of the rapid worldwide increase in what we might call PISA abuse the tests should carry a similar warning to what appears on cigarette packets in Britain: 'This product may damage your nation's educational health.' Further, even though government analysts ought to have the wit to caution against the kind of response that has now become so familiar, OECD must accept some of the responsibility because it presents so much of the data in league table form.

The year of PISA 2012, I suggest and hope, is the year when we recover a sense of proportion and perspective about what these international achievement surveys tell us and what - perhaps even more important - they do not and cannot tell us. We must also recognise the evidential bias, methodological weakness and simplistic thinking in those Type II cause-effect comparative data extrapolations which have had far greater political influence than they deserve. We must much more fully exploit the riches of those other kinds of comparative evidence which in recent decades governments have ignored. We must develop an account of educational progress and performance which doesn't arbitrarily restrict itself to what can be measured. And we must replace the rampantly supremacist or narrowly nationalist view of education by a vision which is more in tune with the true complexities of globalisation, with the perilous condition of our world and with the needs of the world's children.

REFERENCES

Alexander R.J., Broadfoot P., Phillips D. (ed) (1999) *Learning from Comparing: new directions in comparative educational research*, Volume I: *Contexts, Classrooms and Outcomes*, Oxford: Symposium Books.

Alexander, R.J., Osborn, M., Phillips, D. (ed) (2000) *Learning from Comparing: new directions in comparative educational research*, Volume II: *Policy, Professionals and Development*, Oxford: Symposium Books.

Alexander, R.J. (2001) *Culture and Pedagogy: international comparisons in primary education*, Oxford: Blackwell.

Alexander, R.J. (2008a) *Essays on Pedagogy*, Abingdon: Routledge.

Alexander, R.J. (2008b) *Towards Dialogic Teaching: rethinking classroom talk*, York: Dialogos.

Alexander, R.J. (2008c) *Education for All, the Quality Imperative and the Problem of Pedagogy* (CREATE Research Monograph 20), Universities of Sussex and London, with the Department for International Development: Consortium for Research on Educational Access, Transitions and Equity. http://www.create-rpc.org/pdf_documents/PTA20.pdf.

Alexander, R.J. (2010a) ‘“World class schools” – noble aspiration or globalised hokum?’, BAICE Presidential Address at the 10th UKFIET Conference on Education and Development, *Compare* 40(6), 801-817.

Alexander, R.J. (ed) (2010) *Children, their World, their Education: final report and recommendations of the Cambridge Primary Review*, Abingdon: Routledge.

Alexander, R.J., Doddington, C., Gray, J., Hargreaves, L. and Kershner, R. (ed) (2010) *The Cambridge Primary Review Research Surveys*, Abingdon: Routledge.

Barber, M. and Mourshed, M. (2007) *How the World’s Best-Performing School Systems Come Out on Top*, Dubai: McKinsey and Company.

Barber, M., Whelan, F., Clark, M. (2011) *Capturing the Leadership Premium: how the world’s top school systems are building leadership capacity for the future*, Dubai: McKinsey and Company.

Benavot, A. (2008) ‘The organisation of school knowledge: official curricula in global perspective’, in Resnick, J. (ed) *The Production of Knowledge in the Global Era*, Rotterdam and Taipei: Sense Publishers.

Benavot, A., Cha, Y-K., Kames, D., Meyer, J.W., Wong, S-Y (1991), ‘Knowledge for the masses: world models and national curricula, 1920-1986’, *American Sociological Review*, 56.

Bernstein, B. (1970) ‘Education cannot compensate for society’, *New Society*, 26 February, 344-7.

Bloom, K. (2010) ‘Shanghai kids the world’s smartest’, *Daily Maverick* (South Africa), 9 December.

Boyer, E.L. (1983) *High School: a report on secondary education in America*, New York: Harper & Row, p. 6, quoted in H.J. Noah ‘The use and abuse of comparative education’, in Altbach, P.G., Kelly, G.P. (ed) *New Approaches to Comparative Education*, Chicago: University of Chicago Press.

Coffield, F. (2012) 'Why the McKinsey reports will not improve school systems', *Journal of Education Policy*, 27(1), 131-149.

Department for Education (2010) *Remit for the Review of the National Curriculum in England*, London: DfE.

Department for Education (2011a) *The Framework for the National Curriculum: a report by the Expert Panel for the National Curriculum Review*, London: DfE.

Department for Education (2011b) *Review of the National Curriculum in England: report on subject breadth in international jurisdictions*, London: DfE.

Department for Education (2012) 'National Curriculum Review', letter from the Secretary of State accompanying the draft programmes of study for English, mathematics and science, 11 June.

Eisner, E.W. (1997) *The Enlightened Eye: qualitative inquiry and the enhancement of educational practice*, (2nd edition), New York: Merrill Publishing Company, 199.

Feniger, Y. and Lefstein, A. (2012) 'Education policy, culture and maths achievement: immigration as a "natural experiment"', unpublished article.

Fredriksson, P. (2006) 'What is so special about education in Finland?' Paper for the EU Presidency Conference, Helsinki.

Gage, N. (1978) *The Scientific Basis of the Art of Teaching*, New York: Teacher's College Press.

Gray, J., Galton, M., McLaughlin, C., Clarke, B. and Symonds, J. (2011) *The Supportive School: wellbeing and the young adolescent*, Newcastle upon Tyne: Cambridge Scholars Publishing.

Hargreaves, D.H. (2011) *A Self-Improving School System in International Context*, Nottingham: National College for School Leadership.

Hartman, B. (2010) 'PISA test results show colossal failure of Israel schools', *Jerusalem Post*, 9 December.

Heath, S.B. and Street, B.V. (2008) *Ethnography: approaches to language and literacy research*, London and New York: Routledge and Teachers College Press.

Jerrim, J. (2011) *England's 'plummeting' PISA test scores between 2000 and 2009: is the performance of our secondary school pupils really in decline?* DoQSS Working Paper 11-09, Institute of Education, University of London.

Klostermann R and Wolff K (2010) *Die Bild*, <http://www.bild.de/regional/hamburg/pisa/was-machen-die-sachsen-besser-6509694.bild.html> (accessed July 2012)

Li, J. (2012) *Cultural Foundations of Learning: east and west*, New York: Cambridge University Press.

Loveys, K. (2010) in *Mail Online on Sunday*, 8 December, <http://www.dailymail.co.uk/news/article-1336410/OECD-condemns-British-education-inferior-Estonias.html>

Lyytinen, H.K. (2002) 'Why are Finnish students doing so well in PISA?' Paris: OECD.

Mourshed, M., Chijioko, C., Barber, M. (2010) *How the World's Most Improved School Systems Keep Getting Better*, Dubai: McKinsey and Company.

National Research Council (2003) *Understanding Others, Educating Ourselves: getting more from international comparative studies in education* (C.Chabott and E.J.Elliott, editors), Washington DC: The National Academies Press, <http://www.nap.edu/openbook.php?isbn=0309088550>

- News Wales* (2010) Anonymous article, 7 December, <http://www.newswales.co.uk/index.cfm?section=Education&F=1&id=20259>
- Oates, T. (2010) *Could Do Better? Using international comparisons to refine the national curriculum in England*, Cambridge Assessment.
- OECD (2010a) *PISA 2009 Results: learning trends*, Paris: OECD.
- OECD (2010b) *PISA 2009 Results: what makes a school successful?* Paris: OECD.
- OECD (2010c) *PISA 2009 Results. Overcoming social background: equity in learning opportunities and outcomes (Volume II)*, Paris: OECD
- OECD (2010d) *PISA 2009 Results: executive summary*, Paris: OECD.
- OECD (2011a) *PISA 2009 Results: quality time for students – learning in and out of school*, Paris OECD.
- OECD (2011b) *PISA 2009 Results: against the odds*, Paris: OECD.
- OECD (2012) 'What PISA assesses', http://www.pisa.oecd.org/pages/0,3417,en_32252351_32235918_1_1_1_1_1_1,00.html (accessed February 2012).
- Nichols, S. and Berliner, D.C. (2008) *Collateral Damage: how high-stakes testing corrupts America's schools*, Cambridge MA: Harvard Education Press.
- Osborn, M., Broadfoot, P., McNess, E., Planel, C., Ravn B. and Triggs, P. (2003) *A World of Difference? Comparing learners across Europe*, Maidenhead: Open University Press.
- Popham, W.J., Eisner, E.W., Sullivan, H.J., Tyler, L.L. (1969) *Instructional Objectives* (AERA Monograph Series on Curriculum Evaluation 3), Chicago: Rand McNally.
- President's Committee on the Arts and the Humanities (2011) *Reinvesting in Arts Education: winning America's future through creative schools*, Washington, DC: President's Committee on the Arts and the Humanities.
- Raffe, D. (2011) 'Policy borrowing or policy learning? How (not) to improve education systems', *CES Briefing*, 57.
- Ravitch, D. (2010) *The Death and Life of the Great American School System*, New York: Basic Books.
- Ravitch, D. (2012) *Diane Ravitch's Blog*, <http://dianeravitch.net/>
- Resnick, L.B., Asterhan, C., Clarke, C. and Hofkens, T. (eds) (2012, forthcoming) *Socializing Intelligence* [papers from the September 2011 AERA conference 'Socializing intelligence through academic talk and dialogue' held at the University of Pittsburgh], Washington DC: AERA.
- Reynolds, D. and Farrell, S. (1996) *Worlds Apart? A review of international surveys of educational achievement involving England*, London: Ofsted.
- Robinson, P. (1999) 'The tyranny of league tables: international comparisons of educational attainment and economic performance'. In R.J.Alexander, P.Broadfoot, D.Phillips (ed) (1999) *Learning from Comparing: new directions in comparative educational research. Volume I: Contexts, Classrooms and Outcomes*. Oxford: Symposium Books, 217-235.
- Rollot, C. and de Vergès, M. (2007) in *Le Monde*, 4 December,

http://www.lemonde.fr/societe/article/2007/12/04/la-france-paralysee-devant-ses-mauvais-resultats-scolaires_985481_3224.html (accessed July 2012)

Sadler, M. [1900] 'How can we learn anything of practical value from the study of foreign systems of education?' In J.H.Higginson (ed) *Selections from Michael Sadler: studies in world citizenship*. Liverpool: DeJall and Meyorre.

Sahlberg, P. (2011), *Finnish Lessons: what can the world learn from educational change in Finland?* Teachers College Press

Schleicher, A. (2009) 'Wie einer auszog, das Lernen zu messen: was den PISA-Erfinder Andreas Schleicher austreibt', *Geowissen* 44/2009, <http://www.buecherkiste.org/-epaper-geowissen-442009-die-ideale-schule> (accessed May 2012).

Schleicher, A. and Stewart, V. (2008) 'Learning from world-class schools', *Educational Leadership* 66(8).

Tobin, J.J., Hsueh, Y., Karasawa, M. (2009) *Preschool in Three Cultures Revisited*, Chicago: University of Chicago Press.

Tobin, J.J., Wu, D.Y.H., Davidson, D.H. (1989) *Preschool in Three Cultures: Japan, China and the United States*, New Haven: Yale University Press.

Tymms, P. and Merrell C. (2010) 'Standards and quality in English primary schools over time: the national evidence', in Alexander, R.J., Doddington, C., Gray, J., Hargreaves, L. and Kershner, R. (ed), *The Cambridge Primary Review Research Surveys*, Routledge, 435-460.

UNICEF (2007) *Child Poverty in Perspective: an overview of child well-being in rich countries*, Innocenti Report Card 7, Florence: UNICEF Innocenti Research Centre.

UNICEF (2011) *Child Well-being in the UK, Spain and Sweden: the role of inequality and materialism*, London: UNICEF UK.

Whetton, C., Ruddock, G. and Twist, L. (2010) 'Standards in English primary education: the international evidence', in Alexander, R.J., Doddington, C., Gray, J., Hargreaves, L. and Kershner, R. (ed) *The Cambridge Primary Review Research Surveys*, Routledge, 461-483.

Wilkinson, R. and Pickett, K. (2010). *The Spirit Level: why equality is better for everyone*, London: Penguin.

Wolf, A. (2002) *Does Education Matter? Myths about education and economic growth*, London: Penguin Books.

Young, T. (2010) 'British schoolchildren now ranked 23rd in the world, down from 2012th in 2000', *Daily Telegraph*, 7 December.