

Alternative models for analysing and representing countries' performance in PISA

Peter Mortimore

**Paper commissioned by
Education International Research Institute**

Brussels, 2009



**Education International
Internationale de l'Éducation
Internacional de la Educación**

Alternative models for analysing and representing countries' performance in PISA

Summary

This independent report has been commissioned by Education International. Its purpose is to raise questions about the current form and focus of PISA and, where possible, to suggest how these might be improved.

It is recognised that PISA offers a great deal of information to all those involved with schooling and school systems. It provides a treasure trove of valid and reliable data.

At the same time, PISA also suffers some limitations: it assesses a very limited amount of what is taught in schools; it can adopt only a cross-sectional design; it ignores the role and contribution of teachers; and the way its results are presented – in some, at least, of its tables – encourages a superficial, 'league table' reading of what should be a more interesting but essentially more complex picture.

There are few easy solutions to some of these problems. But, with goodwill from the PISA Team, improvements could be made. This Report recommends consideration of a number of changes. It especially commends:

- shifting the aims of PISA from a snapshot of national achievements to a more nuanced interpretation of countries' strengths and weaknesses in their development of lifelong learning amongst their populations
- refocusing on how schools and school systems could promote achievement and increase the equity of their educational outcomes
- widening the basis of assessment through inclusion of knowledge from, say, geography and social science domains
- involving teachers in the design and development phases of future PISA cycles
- including information from teachers to enrich the contextual power of the data
- extending the methodology to include a longitudinal element
- reconfiguring to an alphabetical order the minority of tables currently presented in rank positions.

Introduction

The Programme for International Student Assessment (PISA) is a large-scale international project undertaken by the Organisation for Economic Cooperation and Development (OECD) in conjunction with over fifty national governments. Launched in 1997, it involves triennial cycles of tests in reading literacy, mathematical and scientific competence (and, in 2003, problem solving) of national samples of 15 year-old students in order to consider how well they "are prepared to meet the challenges of today's knowledge society"¹. According to PISA, the Programme provides the following outcomes:

- *profile of knowledge and skills among 15-year-olds.*
- *contextual indicators relating results to student and school characteristics.*
- *knowledge base for policy analysis and research.*
- *trend indicators showing how results change over time².*

In 2000 the initial survey focused on reading literacy, three years later its focus was on mathematical competence and in 2006 it was on scientific competence. In each case, less detailed information was provided for the two other areas. Each survey has resulted in a detailed report. PISA also uses questionnaires to collect information about students' homes and schools. The latest report included information drawn from a parent questionnaire administered in a sample of countries. According to PISA, information from the questionnaires is used in conjunction with the achievement data to explore:

- *differences between countries in the relationships between student-level factors (such as gender and socio-economic background) and achievement;*
- *differences in the relationships between school-level factors and achievement across countries;*
- *differences in the proportion of variation in achievement between (rather than within) schools, and differences in this value across countries;*
- *differences between countries in the extent to which schools moderate or increase the effects of individual-level student factors and student achievement;*
- *differences in education systems and national context that are related to differences in student achievement across countries;*
- *through links to PISA 2000 and PISA 2003, changes in any or all of these relationships over time³.*

PISA offers participating national governments a range of evidence with which to monitor the performance of its own educational systems. It provides data which can be used by governments and their electorates - as well as teachers and other education professionals - to debate the strengths and weaknesses of their schools and their education systems in relation to those of other countries. By providing succinct discussions on the "*implications for policy*" of all major findings, PISA also seeks "...to provide a new basis for policy dialogue and for collaboration in defining and implementing educational goals..."⁴.

PISA currently misses, however, many links to life-long learning. Given the common agreement that a 21st century productive life – and indeed a happy one – is dependent upon continuous learning, this is a serious omission. Possibly instruments such as the *Index of future-oriented motivation to learn science* - used in the 2006 survey⁵ - could be built upon in order to help address this need?

Education International (EI) and PISA

According to an EI publication "*the impact of PISA through the media is undeniable*"⁶. It would be foolish, therefore, for education unions to ignore it. But there are also other reasons to study PISA: the wealth of information produced about student achievement and the correlations with students' socio-economic backgrounds and school and educational system characteristics can add to our knowledge about how schools and school systems can promote achievement and increase the equity of educational outcomes.

EI's main concern is that PISA data should be used more constructively. It is worried that the emphasis on the assessment of reading literacy, mathematics and science devalues all the other work of schools, including the study of other subjects. Moreover, the fascination of the media with crude league tables, reporting the performance of individual countries, encourages superficial debate of what are more complex issues. Furthermore, an EI review of the 2003 PISA survey showed that "education ministers and their advisers... try to put 'spin' on the data to suit their own purposes"⁷.

This independent Report has been commissioned by EI to study these issues, and other criticisms of the PISA approach, in order to consider whether alternative methods of presenting data are more likely to lead to serious debate about the quality and equity of education methods and systems. Its perspective is that of a ‘data user’ who works with both academics and school practitioners and who, in the light of many years of school research⁸, uses the term ‘effectiveness’ to indicate a state where all students are given access to the maximum quality of schooling - in pedagogical, rather than simply in economic, terms.

General criticisms of PISA

There have been many critiques covering the conceptual, methodological and political aspects of the Programme including: Prais, 2003, 2004⁹; Smithers, 2004¹⁰; Goldstein, 2004¹¹; Burckhardt, 2005¹²; Wuttke, 2005¹³; Hopmann and Brinek, 2007¹⁴ (which includes a collection of critical papers by European researchers); and, most recently, by Grek, 2009¹⁵.

Drawing on these critiques, ten important issues emerge.

1. *Cultural differences*

For any programme of assessment the challenge of having to treat students from over fifty nationalities in a common manner is daunting. Students from a range of cultural backgrounds may react in different ways to common questions and even to a common formal testing situation. The problem may reduce over time as the Programme continues and the expectations of teachers and younger students develop. Clearly it should not be the intention of test designers to iron out such cultural differences (which may be crucially important to the identities of different peoples). It is vital, however, that the existence of such differences be borne in mind when interpreting PISA outcomes.

2. *Translations*

There will also be differences in the way ideas can be translated. It is likely that some languages are more difficult than others; those with more regular grammatical constructions, for instance, may be less likely to generate reading or spelling problems. Furthermore, it must be recognised that some students will not be tested in their native language. This will apply to many first generation immigrants but also to students in countries such as Luxembourg where they will be tested in either French or German. PISA addresses the issues of translation in the 2006 Technical Report¹⁶, detailing the steps that have been taken - including both translation and back translation of questions and test items - to overcome at least some of these problems.

3. *Sampling*

With any survey - as opposed to a census which includes everyone of a certain age - there will always be questions about the representativeness of the selected sample. There are dangers that some countries will endeavour to increase the proportion of the most able students and decrease the proportion of those deemed to be the least able. But students with learning difficulties and those lacking fluency with the native language also need to be included in a correct proportion. A number of criticisms have focused on this issue, including *Disappearing Students: PISA and Students with Disabilities* by Bernadette Hörmann of the University of Vienna¹⁷.

The PISA Guidelines address this topic in considerable detail. Each country’s sample is examined carefully and measured against agreed criteria. There are instances where it has

been decided that the sample was not good enough - as with the rejection of the UK sample in 2003. (As a result, the UK outcomes were not published in the normal way but were demoted to an appendix. Subsequent analysis showed that the UK sample was unrepresentative being, in fact, a more able sample than would have been produced by chance¹⁸.)

4. *Disregard of national curricula*

The emphasis on asking questions which can be answered using common sense rather than knowledge of a particular curriculum has also been challenged. Commentators, such as Thomas Jahnke of the University of Potsdam, argue that this leads politicians and the public to form an impoverished view of the curriculum¹⁹. It has also been questioned by EI which notes that “*PISA does not transmit the total picture of education*”²⁰. Currently, as PISA increasingly is used as the ultimate reference on “quality of education” it shifts the attention of the general public and politicians almost exclusively to the core subjects of mathematics, science and reading. This leaves history, geography, civics, foreign languages, and all the other subjects taught in schools, marginalized.

The PISA approach is to assess young people’s ability to use their knowledge and skills “*in order to meet real-life challenges rather than how well they had mastered a specific school curriculum*”. PISA thus places the emphasis on the mastery of “*processes, the understanding of concepts, and the ability to function in various situations within each domain...reading, mathematics and science*”²¹.

If PISA data are seen merely as a crude estimate of the performance of a nation’s educational system and its attempts to achieve equity, the limitations of the curriculum being assessed may not matter. If, however, the data are considered to be the definitive judgement of national systems, then the results will be damaging. The powerful backwash will have the effect of permanently dividing the curriculum into a core (literacy, mathematics and science that which is assessed by PISA) and a less important periphery of all other subjects.

One way to avoid such a danger is to integrate some content from those subjects which contains knowledge of international value – as did the 2003 PISA work on problem solving. Thus the domains of geography or some aspects of social science might provide suitable topics with which to broaden the current PISA assessments.

5. *Distortion of educational policies*

EI argues that PISA, by focusing on a limited set of skills, has encouraged governments to adopt an overly ‘economic’ approach to education - “*testing pupils, evaluating teachers with reforms inspired by efficiency or competitiveness criteria*”²². Some critics have suggested that the private companies, used by PISA, are simply seeking to increase their share of “*an ever-growing testing market*”²³ or, even more extremely, that “*PISA is a New Public Government outlet of the most neo-liberal kind*”²⁴.

6. *Lack of sufficient involvement of teachers*

PISA does not currently collect data from teachers. Yet, in addition to the assessment, it does collect additional information by way of principal and student questionnaires and, in 2006, it also collected information from a sample of parents in sixteen countries – “*The questionnaire collected information about parents’ investment in their children’s education and their views on science-related issues and careers*”²⁵. It appears somewhat

peculiar, therefore, that those working in classrooms - at the heart of education and who have the most to contribute to the improvement of both quality and equity - are ignored.

Of course, some teachers may not wish to get involved; others may not wish to devote precious time to what they see as a marginal activity. Others, however, may welcome the opportunity to participate and, by so doing, may help fill the space which now exists between the school policies - as expressed by the principal - and the students' views and outcomes.

7. *The use of a cross sectional design*

The design of PISA is cross sectional. Triennially, nationally samples of same-aged students are assessed. Although it is likely that each sample will be generally similar to its forerunner, in some cases, the characteristics of the sample may have changed (due, for example, to an increase or decrease in immigration) and this will make comparisons over time more difficult.

As one critic has noted: "*Observed differences will undoubtedly reflect differences between educational systems but they will also reflect social and other differences that cannot fully be accounted for*"²⁶. Furthermore, to make valid comparisons of the effects of different national systems of education or to imply causality, strictly speaking, it is necessary to have longitudinal data. Goldstein makes the point: "*The (2003) Report claims that literacy level 'has a net direct effect on pre-tax income, on employment, on health'. Such causal effects may indeed exist but they cannot be inferred from cross-sectional studies*"²⁷.

A number of commentators note that PISA is used – particularly by governments and the media - to imply causality, even though it is restricted to drawing on correlations between different measures.

Adopting a longitudinal design for at least part of the survey would strengthen the influence of PISA. It would also, of course, increase its costs.

8. *Modelling data*

Goldstein²⁸ draws attention to the technique used by PISA to decide whether items fitted single dimension scales - item response modelling – which has been criticised by Blum et al "*on the grounds that it can lead to subtle biases and may 'smooth out important country differences'*"²⁹. Goldstein also notes that "*items that worked differently in some countries...were suspected of cultural bias...As a result, some items were excluded.*" Such items are labelled as 'dodgy' in the PISA technical report³⁰.

PISA has adopted some multi-level modelling techniques in order to take account of school differences while allowing for student characteristics - such as gender and social background. This is helpful since it is able to incorporate residual variation in the models.

9. *Effects of league tables on national school systems*

Although PISA data are presented in great detail, and meticulous attention is paid to the inclusion of standard errors and multiple comparisons of mean performance showing which national differences are statistically significant, discussion in the media is almost exclusively in terms of countries' rank positions. The subsequent league tables so dominate the news that much of the careful work carried out by the PISA Team on the

relationship between domains, contextual factors, attitudes and school and system factors is ignored – except by scholars.

The media emphasis on such limited data has encouraged the adoption of league tables in individual countries. According to EI: *“These actions usually lead to the introduction of more measurements, of national testing systems based on the PISA model and methodology, of more scrutinized teaching procedures, and ultimately, to linking teachers’ performance and pay to students’ test scores”*³¹.

The experience from England³², where testing of whole age cohorts using standard achievement tests (Sats) has been made statutory, is that league tables dominate public discussion about schools. Such tables fuel the development of a market economy of school places despite being fundamentally flawed by failing to reflect differences in the intakes to schools³³.

Many parents, however, have been persuaded by the national government’s sponsorship of league tables into thinking that they offer a clear guide to the best schools. Accordingly, they are followed avidly by aspiring parents who wish to do the best for their children. A further problem is that, where there are more applicants than places, the exercise of choice switches from parents to schools. In England there have been cases of false addresses being given in order to gain access to certain schools, resulting in court proceedings against parents.

Even the price of houses is affected by league tables, with properties in the catchment of a highly ranked school attracting a sizeable premium. High ranked schools are usually swamped with applications. Low ranked schools often struggle to achieve sufficient numbers to make them viable.

League tables also tend to depress equity as teachers are encouraged to focus on those students who may help lift the school’s rankings. Those who are likely to be borderline candidates in tests are likely to be given the most help. Students who require special help are less likely to be welcomed as they may use more resources – and do less well in the tests - so pulling down the school’s scores.

Furthermore, it is clear from OECD research that parents who are socially, economically or culturally advantaged are more adept at making choices³⁴. In a system based on league tables, the most sought after school places are very likely to be commandeered by such parents, leaving the remainder of less desirable ones for those with less advantage.

For the last twenty years, life in English schools has been dominated by high stakes testing. Marshall (2004) calculated that, on average, each student sits 100 formal tests or examinations³⁵. Mansell (2007) has described the stultifying effects of so much testing³⁶. The key problem with league tables is not the essential testing as such but the way information derived from the tests is used to impose high stakes accountability - with serious consequences for students, teachers and schools. Testing has a positive part to play in students’ learning through the provision of helpful feedback to students and their teachers. The problem arises because the focus of high stakes testing is not on learning, but accountability³⁷.

Even if governments do not sponsor league tables, the media are likely to do so. PISA, therefore, needs to consider how to present its data in order to limit the opportunities for league tables and to encourage a more thoughtful and more positive approach to its work.

10. Dominance and secrecy

OECD and other international bodies have undoubtedly strengthened their influence as discussion about globalisation has increased amongst governments. Although economic policy remains paramount within OECD, its interest in education “*has taken on an increased importance within that mandate, as it has been reframed as central to national economic competitiveness within an economic human capital framework and linked to an emerging ‘knowledge economy’*”.³⁸ According to Grek, “*OECD had created a niche as a technically highly competent agency for the development of educational indicators and comparative educational performance measures*”³⁹.

Other commentators have argued that because governments’ powers over economic trends has weakened – as the latest worldwide financial upheaval has demonstrated – “*the competitive advantage of nations is frequently redefined in terms of the quality of national education and training systems judged according to international standards*”⁴⁰. Thus PISA, supported by the governments of over 50 countries, with its clear technical skills and its flair for presenting complex information in an accessible way, has become one of the world’s key arbiters of the quality and the equality of school systems.

According to some German critics, however, PISA’s value has been diminished by the reported unwillingness of the members of its German consortia to engage in open debate about its assumptions, its methods and its consequences. In an open website, Hopmann et al state that “*repeated invitations to address these issues in open symposia, or to contribute to this volume (PISA according to PISA), remain unanswered or were turned down*”⁴¹. These commentators list a number of complaints about the German Consortium and its responses to the invitation to scientific dialogue about PISA. If these allegations are true, this is a pity and represents a missed opportunity. It also contrasts markedly with the enthusiastic participation of the head of the PISA Team at a London meeting in 2005⁴².

There is clearly much to be gained by the PISA Team operating a policy of openness and demonstrating its willingness to be challenged. It is also important that academics, who may have worked in the field of assessment and struggled for modest research funding over many years, do not allow feelings of resentment to affect their attitude towards the well-funded, internationally supported PISA Project. There are enormous potential benefits to collaborative work and an excellent opportunity for further study using the full PISA data set made available by the Team. The ‘statlinks’ providing excel spreadsheets for each table is an outstanding innovation.

It is hoped that the PISA Team’s response to this paper will further the collaboration between it and both academics and practitioners working in school systems.

The need for change

The issues most relevant to EI requiring change are:

1. Involving teachers

There are two ways in which the involvement of teachers could be brought about. The first is by much greater use in the development of assessment strategies and the editing of test items. The use of teacher focus groups, for instance, would enable the professional knowledge of teachers to be used at key stages in the development of future cycles of PISA.

The second way is by the PISA Team developing questionnaires for teachers in the sample schools. The data generated by such instruments would supplement views of students and parents and thus provide an enhanced context in which to interpret the results of the assessments.

While accepting that some teachers, and indeed some national unions, may feel ambivalent about taking part in a PISA survey, it would be prudent for the PISA Team to offer this possibility.

Thus, information could be collected about teachers' expectations, attitudes, teaching strategies and professional development plans as well as their views of students. Some of the questions could be adapted from the OECD's TALIS project⁴³, others would need to be formulated in order to complement those being asked of students, principals and parents.

It is important to recognise that such extra involvement may have implications for teachers' working time.

2. Using longitudinal data

The limitations of this design have been spelled out earlier in this Report. There are three main arguments for the use of a longitudinal design:

- measures of progress are better than one-off 'snapshots'
- individual longitudinal data are richer than cross sectional material
- inferences about causality can be made on the basis of longitudinal information.

Statisticians such as Goldstein⁴⁴ have argued for the inclusion of a longitudinal element in future cycles of PISA. The PISA Team should consider adding such a component – which could be on a smaller scale – to complement the work of the cross sectional survey.

3. Ameliorating the use of league tables

It is accepted that, in general, PISA data are well presented. Volume 1 contains a vast amount of information in a variety of mainly 'fit-for-purpose' formats. Volume 2 contains over 175 tables with countries listed in alphabetical order.

The sections of the Report dealing with '*What PISA measures and how*' and '*Performance in PISA: what is measured*' and the Figure 1.2 showing a '*Summary of the assessment areas in PISA 2006*'⁴⁵ are models of clarity. Furthermore, Box 2.2 on page 50, '*Interpreting sample statistics*', deals well with the issues of uncertainty, confidence intervals and statistical significance. It should enable a reader, unused to dealing with statistics, to appreciate the limitations as well as the

strengths of the data. Box 3.2 *Interpreting the PISA indices* is very helpful as is Box 5.1 – dealing with the limitations of the analyses - and Boxes 5.2 – 5.9 which deal with the multilevel modelling.

As with any complex report, data are presented in various formats. These include ***alphabetic lists*** of participating countries ranging from Australia to the United States for the OECD countries and from Argentina to Uruguay for the partner countries; ***ranked lists*** according to performance; ***scattergrams*** illustrating the relationship of two or more variables; and ***multilevel charts*** – illustrating the impact of different ways of modelling the data.

1. Alphabetic lists

These are used for listing outcomes of different scales (Figure 2.13), displaying the values of some of the many indices presented in the Report (Figure 3.4), and information on the response rates. Such lists are also used for the 170 or so detailed tables in the second volume of the Report devoted to data. Thus almost 200 of the 300 or so tables and other charts making up the Report use an alphabetic format.

2. Ranked lists

These are used to present the outcomes of the three main scales: science, reading and mathematics (see figures 2.11b, 6.8a and 6.20a); the figures showing the percentage of students at proficiency levels in the three tests (see figures 2.11a, 6.1 and 6.19); and for some of the separate subscales (eg. figure 2.24e).

Tables showing a variety of variables – including most (but not all) of the indices which have been constructed - also use a ranked order: sometimes ascending (see figure 5.17) but, more often, descending (see figure 3.14).

3. Scattergrams

The relationships between variables are illustrated with scattergrams. (For example, figure 2.1a shows the relationship of the outcome of the science scale with a measure of the gross domestic product whilst figure 6.10 illustrates the supremacy of female reading scores.

4. Multilevel charts

The Report contains a series of analyses which test different models in order to explore the combined effects of different contextual variables on the test scores. These analyses are illustrated with a set of charts showing the changes in relationships and the strength of associations (see boxes 5.1 -5.9).

The most common way of presenting data, however, is through the use of alphabetic lists. The problem is that those charts the media are most likely to be interested in are not presented in this format but, instead, appear in ranked orders. Thus a ‘league table’ style appears to be promoted by the PISA Team.

This is misleading because the Report is meticulous in its treatment of data and any fair reading of its content would show that the overall order of countries on the three main tests, whilst being of interest, is only a small part of a bigger picture relating test

scores to a host of conditions, policies and other factors. For instance, the main purpose of Figure 2.11b on pages 56/57, which shows multiple comparisons of mean performance on the science scale, is to illustrate that, although data appear in ranked lists, statistically significant differences only occur between the scores of some of the countries. Thus, whilst both Finland and Kyrgyzstan have scores which are significantly different (one in a positive and one in a negative direction) to each of the other participating countries, all the others have more mixed outcomes. In the case of Denmark, for instance, 25 countries have significantly lower scores, 20 have significantly higher ones and 11 are not significantly different.

Figure 6.8a on pages 296/297 does the same for reading. Countries are ranked from the highest score (Korea, 556) to that with the lowest (Kyrgyzstan, 285). In this case, both Korea and Kyrgyzstan have scores which are significantly different to all other countries. In contrast, the mean score for Croatia is significantly higher than that of 24 other countries and significantly lower than another 24 - with seven countries having scores that are not significantly different.

In the same way figure 6.20a, on pages 316/317 presents data in ranked lists for mathematics. In contrast, figure 2.13, which presents data comparing the overall science score with each of its seven scales, lists countries in alphabetical order. It would have been possible to keep to the league table ranking but, wisely, this was not done.

Using alphabetical ordered data for all lists

Surely, it makes sense to use alphabetical data for all lists of countries? Such a policy may not prevent the media from constructing their own league tables, but it would signal that the PISA Team, on principle, does not endorse such an approach and considers it an inappropriate way to display such complex data.

This principled position would need to be stated in the Foreword to the next PISA Report. Currently the Foreword includes the prevailing rationale for comparative international assessments, the short history of the Project, a note of its innovatory features and a list of those responsible for its governance and execution. It could also include prominent ‘health warnings’ about the use of data and a brief argument as to why league tables are considered inappropriate and how their use detracts from the value of PISA. Similar messages could be repeated – suitably phrased – in Box 1.1 *Key features of PISA 2006* and in the Reader’s Guide which currently appears on page 26.

Furthermore, appropriate and prominent ‘health warnings’ should be included in all press releases.

Some of the tables and charts could also be reconfigured so as to encourage a deeper reading of the data. For instance:

- grouping in discrete categories (e.g. using quarters or top, middle and bottom thirds or half standard deviations or other suitable categories)
- adopting more measures such as the difference between each country’s 95th and 5th percentiles scores on different measures
- mixing outcomes with more contextual and descriptive data.

PISA is a remarkable collaborative project. It would be most unfortunate if, through its use of tables of national results presented in rank order, it became associated with the superficial – and often misleading – approach of some of the media. As has been noted, the majority of PISA tables are currently presented in alphabetical order; a small change in the others would lead not only to greater consistency but also to a more appropriate method of data presentation.

Recommendations and conclusions

PISA is an OECD triumph. It provides governments, academics and voters with rich information about their education systems. It has also created numerous starting points for national research projects.

The Programme could be further enhanced by:

- shifting the aims of PISA from a snapshot of national achievements to a more nuanced interpretation of countries' strengths and weaknesses in their development of lifelong learning
- refocusing on how schools and school systems could promote achievement and increase the equity of their educational outcomes
- widening the basis of assessment through inclusion of knowledge from geography and social science domains
- involving teachers in the design and development phases of the cycles
- including information from teachers to enrich the context of the data
- extending the methodology to include a longitudinal element
- reconfiguring to an alphabetical order the minority of tables currently presented in rank positions.

There would be many benefits to these changes in PISA. Amongst these would be greater support for – and use of – PISA data by teachers and the development of a more sophisticated public understanding of schooling which would be less dependent upon sensational claims. Such an understanding would encompass a broader picture of education focused on issues of quality and equity with many consequent implications for governments' policies and practitioners' responses. These are worthwhile prizes.

Notes

- ¹ PISA 2006 Technical Report page 20 –ISBN 978-92-64-04808-9.
- ² PISA 2006 Technical Report page 22 –ISBN 978-92-64-04808-9.
- ³ PISA 2006 Technical Report page 20/21 –ISBN 978-92-64-04808-9.
- ⁴ PISA 2006 Science Competencies for Tomorrow’s World Volume 1 Analysis. Page 3.
- ⁵ PISA 2006 Science Competencies for Tomorrow’s World Volume 1 Analysis. Page 149.
- ⁶ Education International (2007) *Testing, Ranking, Reforming: Impact of PISA 2006 on The Education Policy Debate. Paragraph 12.*
- ⁷ Education International (2007) *Testing, Ranking, Reforming: Impact of PISA 2006 on The Education Policy Debate. Paragraph 13.*
- ⁸ Mortimore. P (1998) *The Road to Improvement: Reflections on School Effectiveness.* Lisse: Swets & Zeitlinger.
- ⁹ Prais. S (2003) Cautions on OECD’s Recent Educational Survey (PISA) *Oxford Review of education* 29,2, 139-163; Rejoinder to OECD’s Response (2004) *Oxford Review of Education* 30, 4, 569-573.
- ¹⁰ Smithers. A (2004) *England’s Education: what can be learned by comparing countries?* University of Liverpool Centre for Education and Employment Research...
- ¹¹ Goldstein. H (2004) International comparisons of student attainment: some issues arising from the PISA study. *Assessment in Education* 11, 3, 319 – 330.
- ¹² Burkhardt. H (2005) *Assessment Instruments and Learning Goals: PISA, TIMSS and other tests* Royal Statistical Society Meeting. London.
- ¹³ Wuttke, J. 2006. Fehler, Verzerrungen, Unsicherheiten in der PISA Auswertung [Mistakes, biases, uncertainties in the PISA analysis]. In *PISA & Co: Kritik eines Programms* [PISA & Co: Critique of a programme], ed. T. Jahnke, and W. Meyerhöfer, 101–54. Hildesheim/ Berlin: Franzbecker.
- ¹⁴ Hopmann. S, Brinek. G u. Retzl, M (2007) *PISA zufolge PISA Hält PISA, was es verspricht?* Wien: Lit-Verlag Schul- und Bildungsforschung, University of Vienna.
- ¹⁵ Grek. S (2009) Governing by numbers: the PISA ‘effect’ in Europe. *Journal of Education Policy* 24. 2. 23-3.
- ¹⁶ PISA 2006 Technical Report pages 89/90 ISBN 978-92-64-04808-9.
- ¹⁷ Hörmann. B (2009) Disappearing Students: PISA and students with disability in S. Hopmann and G. Brinek *PISA according to PISA* University of Vienna. Wien Lit-Verlag 2007 (Reihe Schulpädagogik und Pädagogische Psychologie, Bd.6) <http://bildungswissenschaft.univie.ac.at/fe2/projekte/pisa-zufolge-pisa/>
- ¹⁸ Micklewright. J and Schnepf. S. V. (2006) *Response Bias in England in PISA 2000 and 2003.* Department for Education and Skills, Research Report 771. [On-line article](#), (ISBN: 1 84478 764 8).
- ¹⁹ Jahnke. T (2006) in S. Hopmann and G. Brinek *PISA according to PISA* University of Vienna. Wien Lit-Verlag 2007 (Reihe Schulpädagogik und Pädagogische Psychologie, Bd.6).
- ²⁰ Education International (2007) *Testing, Ranking, Reforming: Impact of PISA 2006 on The Education Policy Debate. Paragraph 14.*

-
- ²¹ PISA 2006 Technical Report page 22.
- ²² Education International (2007) *Testing, Ranking, Reforming: Impact of PISA 2006 on The Education Policy Debate*. Paragraph 25.
- ²³ Bracey. G.W. (2005) Research: Put out over PISA – Assessing for assessment. *Phi Delta Kappan* 86, 10, 797.
- ²⁴ Huisken, F (2005) Der ‘PISA-shock’ und und seine Bewältigung – Wieviel Dummheit braucht/verträgt die republik? (VSA-Verlag) Hamburg.
- ²⁵ OECD (2007) *PISA 2006: Science competencies for tomorrow’s world*. Paris: OECD.
- ²⁶ Goldstein, *ibid*.
- ²⁷ Goldstein, *ibid*.
- ²⁸ Goldstein, *ibid*.
- ²⁹ Blum, A., Goldstein, H. & Guerin-Pace, F. (2001) International adult literacy survey (IALS): an analysis of international comparisons of adult literacy, *Assessment in Education*, 8, 2, 225-246.
- ³⁰ See, for instance, Report 3b: international summary of dodgy items on page 152.
- ³¹ Education International (2009) PISA: Is testing dangerous? <http://www.ei-ie.org/en/article>
- ³² Mortimore. P (2009) School League Tables and Testing. An interview by John Graham. *Professional Voice*, 6, 3, 57 – 61.
- ³³ Mortimore. P (2008) *Lies, damned lies and Sats-based league tables* Education Guardian 05.08.08.
- ³⁴ OECD (2007) *No more failures: ten steps to equity in education*. Paris: OECD.
- ³⁵ Marshall. B (2004) Testing, testing, testing. In T. Wragg (Ed) *Letters to the Prime Minister*. London: New Vision Group.
- ³⁶ Mansell, Warwick (2007) *Education by Numbers*. London: Politico’s.
- ³⁷ See extensive discussion of this point in House of Commons, Children, Schools and Families Select Committee *Testing and Assessment*. Third Report of Session 2007 – 08.
- ³⁸ Grek, *ibid*.
- ³⁹ Grek, *ibid*.
- ⁴⁰ Brown, P., A.H. Halsey, H. Lauder, and A. Stuart Wells. 1997. The transformation of education and society: An introduction. In *Education culture economy society*, ed. A. Halsey, H. Lauder, P. Brown, and A. Stuart Wells. Oxford: Oxford University Press.
- ⁴¹ Hopmann et al, *ibid*.
- ⁴² Department for Children, Schools and Families - Seminar 21 March 2005 at Church House, London.
- ⁴³ www.oecd.org/edu/talis

⁴⁴ Goldstein, *ibid.*

⁴⁵ OECD (2006) *ibid* page 20.

Appendix 1 – Longitudinal studies

The full Report recommended “extending the methodology to include a longitudinal element” on the grounds that:

- measures of progress are better than one-off ‘snapshots’
- individual longitudinal data are richer than cross sectional material
- inferences about causality can be made on the basis of longitudinal information.

If the PISA Governing council is sympathetic to this argument, it may wish to consider two different ways of incorporating a longitudinal approach into its future work.

1. Collection of data from pre-PISA students

This could take place when students were 12 or so years old (in some countries the issue would be complicated by changes of phases of schooling) and were attending the schools in which PISA tests would later be completed. This would provide the advantage that the principals and teachers would already be familiar with PISA protocols and testing procedures. The two phases of testing could be undertaken in parallel at the same time thus removing the necessity for two testing sessions.

As with all testing, it would be essential that standards of validity and reliability were met. Testing would need to be undertaken in the same three PISA domains. Ideally, test items would sample similar skills to those which are assessed at the later age.

Detailed identification of students (confidential to PISA) would be vital in order for the data to be linked accurately to the ensuing PISA outcomes - at the levels of individuals, schools and countries.

The pre-PISA data would enable detailed estimates to be made of the students’ *progress over three years* (rather than just their attainment at the end of lower secondary schooling). They would also permit analyses of individual, school and national effects to be undertaken. Such analyses, used in conjunction with the contextual information already available, would allow much more sophisticated comparisons of different educational systems to be made.

2. Collection of data from post-PISA students

This could take place in upper secondary schools three years after the PISA tests. Because of the change from lower to upper secondary schools in many countries, post-PISA tests would need to be introduced in schools previously unused to PISA procedures. Furthermore, great care would need to be taken in order to reduce sample mortality and to ensure that identification of students was matched in both lower and upper secondary schools.

As with pre-PISA tests, there would be a need for validity and reliability and the domains should remain the same - although it would be expected that the level of performance would be considerably higher. The advantage of testing at this phase of schooling is that it would provide a detailed follow-up of students who had performed well or badly at age 15.

The post-PISA progress of particular groups could be identified and the positive effects of particular school experiences could be investigated. Again such data, if used in conjunction with the contextual information, would be extremely valuable to participating countries.

The collection of longitudinal data at either pre- or post-PISA stages would greatly enhance the value of PISA. Either would provide a better indication of causal relations than the correlations currently employed.

The advantages of choosing the *pre-PISA model* are that the work would be undertaken in the same schools and that identification of students would, therefore, be more reliable. Both phases of testing could also be undertaken at the same time. Because no students would be of an age to leave school, the sample mortality would be much reduced.

Appendix 2 – Bar Charts

The full Report recommended “reconfiguring to an alphabetical order the minority of tables currently presented in rank positions”. It would be easy to do and would match the majority of tables already in this format. What would be less easy to transform are those figures containing bar charts presented in either ascending or descending rank orders. Currently these figures give the reader a pictorial indication of the amount of overall country variation - ranging from that with the highest percentage in the most favourable category to that with the highest percentage in the lowest. These figures are less effective, however, in giving a clear indication of the subdivision into the intervening categories.

If the PISA Governing council is sympathetic to change, it may wish to consider the following alternative technique.

The use of 100% ‘stacked bar’ charts

This could represent the countries (in alphabetical order) in rows with students’ scores divided between the seven categories (from proficiency level below 1 to proficiency level 6). Provided each category contains a data label (indicating the %) the figure would be relatively easy to read, even though the reader would not have the holistic image of ‘best to worst’. (See attached figure which is a new version of part of Figure 6.19).

If different colours could be used for the different countries, the format would be relatively easy to read.

See below, for example:

