

Introduction

0. In the Netherlands an educational war is raging as of about 1996. Innovators and traditionalists fight each other in public debates of which the end is not in sight. The traditionalists are organized into an association named Beter Onderwijs Nederland, BON, (Better Education for the Netherlands) that registered more than 4000 members, mostly teachers in all sectors of the educational system, within one year of its existence. Among other activities BON runs its own site www.beteronderwijsnederland.nl and initiates and supports discussions with their opponents, mostly educational managers and politicians. In so far as discussions are sensitive to outcomes of relevant educational research BON tries to construct a knowledge base for good education. One chapter of this knowledge base critically surveys the outcomes of PISA and TIMSS. Dutch results were until now rather positive, though less so with each successive trial, outcomes that enabled to silence BON-critique on innovations rather effectively by stating: "But internationally our results are good." The following paragraphs are translated from the Dutch text to be found at the BON-site:

www.beteronderwijsnederland.nl/?q=node/1340

All official PISA (2000, 2003, 2006) and TIMSS (1995, 1999, 2003) reports of many sorts, adding to many thousands of pages, are free downloadable at the PISA and TIMSS-sites mentioned in the References. Other texts cited can be consulted at the BON-site.

As critique on PISA and TIMSS increases into an equally unsurveyable quantity the reader is advised to start with publications that rubricate and summarize critique: Smithers (2004); Prais (2003, 2004), the PISA-response of Adams (2003); Jahnke (2006, especially Wuttke's chapter); Haahr (2005); Hagemester (2006); Topping (2003); Bender (2003).

1. The most important and everywhere to be heard critique concerns the validity of the test items for the national curricula. Added to that is the more usual and substantial critique concerning badly constructed and badly translated items (Hagemester, 1999; Baumert's replique, 1999; Bender, 2003, par.3.2). TIMSS, in contrast with PISA, makes use of test items more closely derived from the national curricula as printed in book form. These items try to answer the question: what did they learn? It is quite normal that test items of international surveys come under heavy fire; transnational validity obviously is hard to construct. In a country like the USA, where a Math war between realistic (contextual) and traditional (conceptual) mathematics is raging for many years, even a national curricular validity of items is very difficult to reach and not to find. Of course, even without Math Wars and controversial innovations curricula within countries may differ substantially (Bracey, 2000; Clarke, 2001, par. 3). This being the case, test items always will connect better to curricula of country A than of country B, causing variance in the scores not controlled and for. Positions on the country lists of PISA and TIMSS (see paragraph 2 hereafter) change inexplicably and substantially in time (Bender, in Jahnke, 2006, p.193; Smithers, par. 71 - 74). Sometimes countries with widely different curricula end up *ex aequo* (i.e. Belgium and the Netherlands, TIMSS 2003, De Lange, fig. 5, p. 19); countries with about equal curricula very differently (Flanders and the Walloon provinces of Belgium). Explanations are missing so these country lists are no good. "One number doesn't tell all". Indeed, in the first TIMSS-report this is noticed (TIMSS II, Baumert, J et al., 1997: TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. p. 18 ff.), but in later reports this sensible reticence is lacking.

2. PISA will not have anything to do with transnational validity (Smithers, par. 20) and instead introduces the terms: Literacy (reading, RL; mathematics, ML; science, SL) and "skills for life". For a comparison of PISA and TIMSS-items see Smithers, par. 24 t/m 30. The differences seem to be remarkably small. These new constructs made bad things worse. Even de Lange (2006), chair of the PISA Expert Group Mathematics does not succeed in explaining the precise meaning of math literacy for math, nor for daily life either. The same judgment applies to the repeated and verbose attempts undertaken in the final and national reports. This almost certainly means only one thing. Literacies are competencies (Bender, 2003, par. 3.3 t/m 3.7; Bender, 2004): non-definable, difficult to instruct and untestable learning goals, that, when taken seriously by the teacher, cause all education into a state of weightlessness very fast.

A PISA-test, disconnected from the national curriculum, is not a direct measurement of the quality of that curriculum and will not succeed in comparing the educational systems of countries in this respect. Without curricular validity no country lists and no feedback of results to this curriculum. Not even some gain is made by instructional practice because the test items try to measure intangible competencies.

The positions of the countries on the PISA and TIMSS-lists have but little meaning (Haahr, p. 33 ff.).

This has also been conceded by the authors of the first PISA-report (PISA, 2000, Knowledge and Skills for life, p. 26 en p. 212). Again, as with TIMSS, this standpoint disappeared from later reports without a trace or comment. Neither TIMSS, nor PISA offers explanations for unexpected and shifting positions of nations (Smithers, par. 113). Both surveys lack the theoretic stamina needed for that. (Smithers, par. 106).

3. What *does* PISA measure? Its items resemble the American Scholastic Aptitude Test items and those of the Dutch CITO-test. Indeed, with some other organizations CITO produces PISA-items, an advantage for the Netherlands and a disadvantage for countries that have to deal with translations. Meyerhöfer even remarks that some PISA-items come from Dutch schoolbooks (Jahnke, p. 135.). PISA then seems to measure an unsharp mix of intelligence, knowledge and experience, in other words: common sense (Prais, 2003, p. 141 – 145). Smithers draws attention to the fact that the mean scores for the three literacy's are about equal within each country. He therefore concludes that PISA test general reading skills (Smithers, par. 50; 92 – 95). When PISA-scores are indeed under the influence of intelligence all outcomes must be reconsidered, unless of course one is of the opinion, shared by some educational innovators, that intelligence is a competence that is teachable and learnable in schools. In the many official PISA and TIMSS-reports not a single word is spent on this important question.

4. Naturally the educational systems of the participating countries differ on many more aspects than the content and didactics of the curricula. The outcomes are under the influence of all kinds of educational and cultural conditions in which countries differ (Le Tendre, et al. 2001). To mention but a few: availability of instructional material, the wealth and educational spending of nations, discipline in the classroom, the different importance attributed to subjects, working conditions and salaries of teachers, the status of their profession, class size, the certification of teachers, the social-economic-status of parents, experience with multiple choice questions (not in use in many countries, used in about 30% of the test items), the presence and extent of migrant groups of pupils with language deficits. All these and other not mentioned factors have their effects on teaching and learning and therefore on the outcomes of these international surveys. When controlling for these factors is faulty or even missing, then interpretations and conclusions suffer correspondingly and country lists lose their meaning (Bracey, 2000; Wang, 2001).

5. What follows is a small selection of examples of that happening.

5.1 The sampling of respondents and schools caused unsolved problems in some countries (Smithers, 2004, par. 31 - 42; Prais, 2003, p. 145 – 152). Swedish TIMSS-respondents, for example, are so much older than their German counterparts that the position of the two countries on the list is thereby explained (Bender, in Jahnke, 2006, p. 192). More on this in Wuttke, in Jahnke, 2006, par. 8, p. 114. Between countries there are large differences in percentages pupils not taking the test (i.e. test refusers, drop-outs). Not enough information on these groups was gathered to make sure that the replacements were representative (Prais, 2004, p. 571). Much more critical remarks on sampling problems in Collani (2000, par. 3 en 4) who states that, apart from the Netherlands, fifteen other countries should have been removed from the final PISA 2000 report when the PISA-standards would have been applied (ibid. par. 5). Standards were overrun more often, i.e. concerning the rules for exclusion of participation of special groups of pupils and concerning the rules for handling incomplete answer booklets (Wuttke, in Jahnke, 2006, par. 3 - 5 and 7).

5.2 Differences between country scores can sometimes be explained by causes other than covered by PISA and TIMSS. De Lange (2006, p. 20, TIMSS-table 6) reports that in Singapore, winner of the MathLit competition, pupils spend far more time on these subjects than pupils in other countries. As table 6 shows the factor 'student time spend' neatly predicts the outcomes in Math. The factor was however not controlled for although the data were available.

5.3 Pupils with a language deficit caused by a migrant background score of course lower than is normal. For these lower scores see Smithers, par. 68 -70, table 13. This table also gives the very dissimilar percentages of such respondents for each country: Finland 1,2%; Japan 0,1%; Germany 15,2%; Switzerland 20,7%, Canada 20,6%. Moreover countries apply very dissimilar entrance policies, i.e. selection on earlier schooling of the applicants (Ireland and Canada for instance), or concerning the right of family reunion etc. As a rule emigrants from Asian and East-European countries are better educated than emigrants from other countries. Immigrants in English speaking countries as a rule have smaller language deficits than immigrants in countries like the Netherlands and Germany. This factor has not been handled with due care in the surveys (Hagemeister, 2006, par. 7; also in Jahnke, 2006; Bender, 2005, par. 1.5.).

5.4 The importance of class size is underestimated by PISA due to high PISA-scores of countries like Japan and Korea in combination with big classes in these countries. However, research supplies ample evidence that class size is an important predictor of learning effects and that weak pupils especially profit by a big teacher – pupil ratio. For more on this see Hagemester, 2006, par. 2.

5.5 The intransparency and possible inadequate application of psychometric statistics (i.e. Item Response Theory) is treated and criticized by Prais, (2003, Annex, p. 159), Goldstein (2004) and Von Collani (2001). This point deserves more attention than it gets, like the next one. In the national report of the USA (PISA - Outcomes of Learning, 2002, p. 11) the Reading Literacy country list shows a large middle group of twenty countries, including the Netherlands and Germany, which do not differ significantly from the USA mean score. The leading group counts three countries (Finland, Canada, New-Zealand) and there is a tail group of four. Other national reports, i.e. Germany's, (PISA 2000, Zusammenfassung zentraler Befunde, p. 13) show a middle group of six countries and Germany in the tail group, just because one changed from USA mean score to OECD mean score. Inconsistencies like this one abound in the national reports.

It is clear throughout that the differences between countries are small, just 10% of the total variance in the scores is located between countries. It is quite possible that this partly explains the earlier mentioned unexplained shifts in country positions (par. 1).

5.6 PISA has a not really hidden agenda promoting comprehensive schools and so called equity and equality. Finland and some other high scoring countries have comprehensive schools. Finland also has high quality, high status teachers and an excellent remedial system, among other factors that have a positive influence on learning results. More on this by Naylor (2004). Smithers, (par. 61 - 67; 107 - 110) demonstrates some inadequacies of the PISA measures of equality. The political taste of PISA must remain undiscussed.

TIMSS and PISA scores may offer useful information to individual countries, Prais gives an example of this (2003, p. 144). However they don't lend themselves for comparisons of the quality of national educational systems (Smithers, 111 - 114). Comparing such diversity as offer countries like Finland, Singapore, Germany, USA and Turkey along one scale of country lists is utterly impossible with a few elementary instruments as used by PISA and TIMSS.

<a href=<mailto:wsmitharmelen@gmail.com>>Willem Smit

References

Official PISA-reports are downloadable from <http://www.pisa.oecd.org>

TIMSS-reports from the NCES-site: <http://nces.ed.gov/timss/index.asp>

Adams, R. J. (2003). Response to 'Cautions on OECD's recent educational survey (PISA)'. *Oxford Review of Education*, vol. 29, no. 3, 2003.

Bender, P. (2003) *Die etwas andere Sicht auf die internationalen Vergleichsuntersuchungen TIMSS, PISA und IGLU*.

Bender, P. Die etwas andere Sicht auf den mathematischen Teil der internationalen Vergleichsuntersuchungen PISA sowie TIMSS und IGLU. DMV-Mitteilungen 12-2/2004.

Bender, P. (2005). *Neue Anmerkungen zu alten und neuen PISA-Ergebnissen und Interpretationen*.

Bender; more from Bender at <http://math-www.uni-paderborn.de/~bender>.

Baumert, J. e.a. (1999). *Konzeption und Aussagekraft der TIMSS-Leistungstests*. Response to Hagemester (1999).

Bracey, G. W. (2000). The TIMSS "Final year" study and report: A critique. *Educational Researcher*, Vol. 29, May 2000, p. 4 - 10.

Clarke, D. (2001). *Developments in international comparative research in mathematics education: Problematising cultural explanations*.

Collani, E. von (2001). OECD PISA - An example of stochastic illiteracy? *Economic Quality Control*, vol. 16, no. 2, 227 - 253.

Downes, S. (2005). Understanding PISA. *Turkish Online Journal of Distance Education*. 2005, vol. 6, no. 2, art. 1.

Goldstein, H. (2004). *International comparisons of student attainment: some issues arising from the PISA study*.

Haahr, J. H. e.a. (2005). *Explaining student performance*. Evidence from the international PISA, TIMSS and PIRLS surveys.

- Hagemeister, V. (1999). *Was wurde bei TIMSS erhoben? Eine Analyse der empirischen Basis von TIMSS*.
- Hagemeister, V.; *Kritische Anmerkungen zum Umgang mit den Ergebnissen von PISA* (2006).
- Jahnke, T. & Meyerhöfer, W., Hrsg. (2006). *Pisa & Co*. Franzbecker Verlag, Berlin.
- Lange, J. de. *Mathematical literacy for living from OECD-PISA perspective*. (2006).
- Le Tendre, G. K. et al. (2001). Teacher's work: Institutional isomorphism and cultural variation in the U.S., Germany, and Japan. *Educational Researcher*, vol. 30, no. 6, p.3 - 15.
- Naylor, F. (2004). The Trojan horse within. *Current Concerns*, no.1, 2004.
- Prais, S. J. (2003). Cautions on OECD's recent educational survey (PISA). *Oxford Review of Education*, vol. 29, no. 2, 2003.
- Prais, S. J. (2004). Cautions on OECD's recent educational survey (PISA): Rejoinder to OECD's response. *Oxford Review of Education*, vol. 30, no. 4, 2004.
- Smithers, A. (2004); *England's education. What can be learned by comparing countries?* University of Liverpool.
- Topping, K. e.a. (2003). *Policy and practice implications of PISA 2000*. Report of the International Reading Association, PISA Task Force.
- Wang, J. (2001). TIMSS primary and middle school data: some technical concerns. *Educational Researcher*, vol. 30, no. 6, p. 17 - 21.
- Wuttke, J. (2006) Fehler, Verzerrungen, Unsicherheiten in der PISA-Auswertung. In: Jahnke, T. *PISA & Co*, p. 101 - 154.

(2006)

Willem Smit