WHAT DO STUDIES LIKE PISA MEAN TO THE MATHEMATICS EDUCATION COMMUNITY?

Graham A. Jones

Griffith University, Gold Coast, Australia

In a real sense, PISA 2003 has touched the mathematics education community by stealth rather than by storm. Although PISA brings "baggage" commonly associated with international assessments, it takes some refreshing perspectives especially in the way that it envisions and assesses mathematical literacy. In this panel discussion we *focus on some of the issues associated with PISA: scrutiny of student performance, construct and consequential validity, what makes items difficult for students and the potential impact of PISA on mathematics education research. In selecting these issues we merely begin the debate and open the way for your participation.*

WHAT IS PISA?

The *Programme for International Student Assessment* ([PISA], OECD, 2005) is an international standardized assessment in reading literacy, mathematical literacy, problem-solving literacy and scientific literacy. It started in 1997 when OECD countries began to collaborate in monitoring the outcomes of education and, in particular, assessed the performance of 15-year-old school students according to an agreed framework. Tests have typically been administered to 4,500-10,000 students in each country. The first assessment in 2000 which focused mainly on reading literacy surveyed students in 43 countries while the second assessment in 2003 involved 41 countries and focused mainly on mathematics and problem solving. The third assessment in 2006 will largely emphasize scientific literacy and is expected to include participants from 58 countries. In this panel discussion we will concentrate on PISA 2003 and those aspects of it that deal with mathematical literacy.

THE PISA MATHEMATICAL LITERACY ASSESSMENT

In describing their approach to assessing mathematical performance, PISA documents (e.g., OECD, 2004a) highlight the need for citizens to enjoy personal fulfilment, employment, and full participation in society. Consequently they require that "all adults–not just those aspiring to a scientific career–be mathematically, scientifically, and technologically literate" (p. 37). This key emphasis is manifest in the PISA definition of mathematical literacy: " …an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned, and reflective citizen" (OECD, p. 37; see also Kieran, plenary panel papers).

Reflecting this view of mathematical literacy, PISA documents (e.g., OECD, 2004a) note that real-life problems, for which mathematical knowledge may be useful, seldom appear in the familiar forms characteristic of "school mathematics." The

2005. In Chick, H. L. & Vincent, J. L. (Eds.). *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 1, pp. 71-74. Melbourne: PME.

PISA position in assessing mathematics was therefore designed "to encourage an approach to teaching and learning mathematics that gives strong emphasis to the processes associated with confronting problems in real-world contexts, making these problems amenable to mathematical treatment, using the relevant mathematical knowledge to solve problems, and evaluating the solution in the original problem context" (OECD, 2004a, 38). In essence, mathematical literacy in the PISA sense places a high priority on mathematical problem-solving and even more sharply on *mathematical modelling*.

Although PISA's devotion to mathematical modelling has my unequivocal support, my experience tells me that it is not easy to incorporate effective mathematical modelling problems in a test that has fairly rigid time constraints. In addition, although the term mathematical modelling is relatively new in school mathematics (Swetz & Hartzler, 1991), there are instances of mathematical modelling even in the notorious public examinations of more than 50 years ago. I well remember the following problem in an examination that I took in 1953. It seems to me that it is a genuine modelling problem and it was certainly not a text book problem or a problem that anyone of that era had practised. Moreover, the fact that less than 10% percent of the 15 to 16-year-old students taking the examination solved the problem is both déjà vu and prophetic for those setting the directions for the PISA enterprise.

> In a hemispherical bowl of radius 8 inches with its plane section horizontal stands water to a depth of 3 inches. Through what maximum angle can the bowl be tilted without spilling the water? Give your answer to the nearest degree (University of Queensland, 1953)

Accordingly, even though members of our panel valued the PISA emphasis on real-world problems and mathematical modelling, there was no shortage of issues to debate. In particular, there were issues about the framework, the validity of the assessment, the construction of items, the measurement processes, the conclusions and the interpretations especially interpretations that cast the findings into the realm of an international "league table". Consequently, we faced a problem in selecting which issues to examine. Let me presage the papers of the other panellists by providing an entrée of the issues that reverberated over our internet highways.

WHAT ISSUES DOES PISA RAISE FOR MATHEMATICS EUDCATION?

As the conference theme was *learners and learning* we questioned whether PISA assessment really was designed to support a real-world approach to mathematics teaching and learning. We also raised questions about whether student performance in the PISA assessments mirrored student performance in other mathematics education research on learning and teaching. Although appropriate data was not easily accessible, we wondered what the PISA study told us about patterns of classroom activity in different cultures. Yoshinori Shimizu (plenary panel papers) did examine this from a cultural perspective by scrutinizing Japanese students' responses to some PISA items.

Issues associated with item validity, item authenticity, and item difficulty were consistently part of our discussions. The "triangular park problem" (see Williams, plenary panel papers) was hotly debated and members of the team even spent considerable time looking for triangular parks or car parks. This was part of our conversation on *real world* or *authentic assessment* and this issue is taken up further by Julian Williams under the broader topic of construct validity. Carolyn Kieran (see plenary panel papers) takes up the issue of "what makes items difficult for students?" She observes that the *difficulty levels* of some PISA items are problematic and raises doubts about how much we know about what students find difficult in certain mathematical tasks.

The politics of international assessment studies like PISA (OECD, 2004a) and *Trends in International Mathematics and Science Study* ([TIMSS], Mullis et al., 2004) were high on our debate list. Not only do these debates raise highly volatile issues and national recriminations, they also generate profound questions for those countries that are doing well and for those who are not. In addition to issues that focus specifically on the international league, assessment studies like PISA produce a range of related debates about factors such as gender, ethnicity, socio-economic status, systemic characteristics, approaches to learning, student characteristics and attitudes, and of course fiscal support (OECD, 2004b). Julian Williams (see plenary panel papers) tackles a number of these political issues especially those related to accountability: managing targets, dealing with league tables, and performance-related reviews.

There was considerable interest in discussing the impact of international assessment studies on mathematics education research. At the forefront of such issues is the question: What does PISA say to researchers interested in assessment research? Yoshinori Shimizu (see plenary panel papers) will talk about this more specifically as he refers to the benefits that can be gleaned by researchers through an examination of PISA's and TIMSS's theoretical frameworks, methodologies, and findings. For example, he notes that the detailed item scales and maps in PISA will enable researchers to perform a secondary analysis of students' thinking and accordingly gain a deeper understanding of learners and learning. Michael Neubrand (see plenary panel papers) also looks at the potential of PISA to stimulate research in mathematics education. He focuses on the structure of mathematical achievement especially in the way that PISA conceptualizes achievement through the aegis of a *mathematical literacy* framework. This gives rise to an interesting dialogue with respect to both individual and systemic (collective) competencies in mathematics and how they can be measured. There are of course other important questions such as "What do studies like PISA say to mathematics education researchers about methodological issues such as qualitative versus quantitative research?" Although this particular question is not directly addressed, the panel refers frequently to methodological issues and as such issues a challenge to the participants for further engagement and debate.

CONCLUDING COMMENTS

I believe that this panel discussion is most timely as I am not convinced that mathematics educators are as cognizant as they might be about the impact of the burgeoning industry that encompasses international studies like PISA (OECD, 2003) and TIMSS (Mullis et al., 2004). Although the build up and dissemination of PISA has been slow to take root in the mathematics education research community, the findings have certainly not gone unnoticed by national and state governments, educational systems, business leaders and parent groups. They know where their nation or their state came in the "league stakes" but they have little understanding of the intent and limitations of such studies. Accordingly, an important aim of this panel is to encourage mathematics education researchers to be more proactive not only in publicly illuminating and auditing research like PISA but also in identifying ways in which PISA can connect with and stimulate their own research. In the words of Sfard (2004, p. 6) we should exploit these special times in mathematics education:

> Confronting the broadly publicized, often disappointing, results of the international measurements of students' achievements, people from different countries started wondering about the possibility of systematic, research-based improvements in mathematics education

References

Mullis, I. V. S., Martin, M.O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

OECD (2004a). *Learning for tomorrow's world: First results from PISA 2003*. Paris: Author.

OECD (2004b). *Messages from PISA 2000*. Paris: Author.

OECD (2005, May 8). *Program for international student assessment* (PISA). Retrieved from http://www.pisa.oecd.org/document/56.

Sfard, A. (2004).There is nothing more practical than good research: On mutual relations between research and practice in mathematics education. In M. Niss (Ed.), *Plenary and regular lectures: Abstracts*. Copenhagen, Denmark: 10th International Congress of Mathematical Education

Swetz, F. & Hartzler, J. S. (1991). *Mathematical modelling in the secondary school curriculum*. Reston, VA: NCTM.

University of Queensland. (1953). *Mathematics B: Junior public examination*. Brisbane, Australia: Author.

FROM A PROFILE TO THE SCRUTINY OF STUDENT PERFORMANCE: EXPORING THE RESEARCH POSSIBILITIES OFFERED BY THE INTERNATIONAL ACHIEVEMENT STUDIES

Yoshinori Shimizu

Faculty of Education, Tokyo Gakugei University

The recent release of two large-scale international comparative studies of students' achievement in mathematics, the OECD-PISA2003 and the TIMSS2003, has the potential to influence educational policy and practice. A careful examination of their findings, theoretical frameworks, and methodologies provides mathematics education researchers with opportunities for exploring research possibilities of learners and learning.

BEYOND THE COMPETITIVE EMPHASIS IN REPORTS

The release of results of the OECD-PISA2003 (Programme for International Student Assessment, OECD, 2004) and the TIMSS2003 (Trends in International Mathematics and Science Study, Mullis, et al., 2004) in December 2004 received huge publicity through the media in Japan. The purposes of international studies such as PISA and TIMSS include providing policy makers with information about the educational system. Policy makers, whose primary interest is in such information like their own country's relative rank among participating countries, welcome a simple profile of student performance. Also, there is a close match between the objectives of PISA, in particular, and the broad economic and labour market policies of host countries. The match naturally invites a lot of public talk on the results of the study with both competitive and evaluative emphasis. This was the case in Japan.

There was one additional large-scale study in 2003 of student performance in mathematics in Japan. In the National Survey of the Implementation of the Curriculum, which has also been released recently (NIER, 2005), the students from grades 5 through 9 (N>450,000) worked on items that are closely aligned with the specific objectives and content of in Japanese mathematics curriculum. TIMSS2003 sought to derive achievement measures based on the common mathematical content as elaborated with specific objectives, whereas PISA2003 was explicitly intended to measure how well 15-years-olds can apply what they have learned in school within real-world contexts. The recent release of these studies should shed light on the new insight into learners and learning from multiple perspectives.

The large-scale studies, conducted internationally or domestically, provide a profile of a population of students from their own perspectives. We need to go beyond competitive emphasis in the reports of such studies to understand more about the profile of students' performance and to explore the possibilities of further research that such studies provide.

2005. In Chick, H. L. & Vincent, J. L. (Eds.). *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 1, pp. 75-78. Melbourne: PME.

In this short article, a few released items of PISA2003 are drawn upon to propose that a careful examination of the findings, the theoretical framework and the methodology used as well, provides mathematics education researchers with opportunities to examine further research questions that might be formulated and addressed.

THE SCRUTINIES NEEDED

One of the distinct characteristics of the PISA2003, having mathematics as the major domain in the recent cycle of the project, is the way in which the results of student performance are described and reported. The mathematics results are reported on four scales relating to the overarching ideas, as well as on an overall mathematics scale. The characteristics of the items as represented in the map, which shows the correspondence between the item and the scale, provide the basis for a substantive interpretation of performance at different levels on the scale.

We can now take a closer look at the profile of students' response to the released items. Even the results of a few released items from PISA2003 suggest possibilities for conducting a secondary analysis and further research studies in order to develop deeper understanding of learners and learning. In particular, such items, or overarching ideas, as follows raise questions for Japanese mathematics educators, in particular, and mathematics education researcher, in general, to consider.

An Illuminating Example: SKATEBOARD

One of the items on which Japanese student performance looks differently from that of their counterparts elsewhere is in Question 1 of the item called SKATEBOARD (OECD, 2004, p.76). This short constructed response item asks the students to find the minimum and the maximum price for self-assembled skateboards using the price list of products given in the stimulus. The item is situated in a *personal* context, belongs to the *quantity* content area, and classified in the *reproduction* competency cluster. The results show that the item has a difficulty of 464 score points when the students answer the question by giving either the minimum or the maximum, which locates it at Level 2 proficiency. On the quantity scale, 74% of all students across the OECD community can perform tasks at least at Level 2. The full credit response has a difficulty of 496 score points, which places it at Level 3 proficiency. On the quantity scale, 53% of all students across the OECD community can perform tasks at least at Level 3.

When we look into the data on the students' response rate in each country, a different picture appears. Japan's mean score was significantly lower than the OECD average for the item (See Table 1) and the pattern in the percentages for students' responses look different from their counterparts in other countries.

Of note among the numbers in Table 1 is the lower percentage of correct responses from Japanese students than from their counterparts, as well as the higher no response rate. Students can find the minimum price by simply adding lower numbers for each part of the skateboard and the maximum price by adding larger numbers.

Country	Full Credit	Partial Credit	No Response	Correct
Australia	74.1	9.3	1.8	78.7
Canada	74.9	9.1	2.0	79.4
Germany	71.7	11.5	5.2	77.5
Japan	54.5	8.0	10.6	58.5
OECD Average	66.7	10.6	4.7	72.0

Table 1: The percentage of students' response for SKATEBOARD, Question1 (An excerpt from National Institute for Educational Policy Research, 2004, p. 102.)

The results suggest that some students, Japanese students, in this case, may be weak in handling multiple numbers where some judgment is required, assuming that they have little trouble in the execution of the addition procedure. We need an explanation with scientific evidence for the results.

Another Example: NUMBER CUBES

Another example comes from the result of the item called NUMBER CUBES (OECD, 2004, p.54). This item asks students to judge whether the rule for making a dice (that the total number of dots on two opposite faces is always seven) applies or not with the given four different shapes to be folded together to form a cube. The item is situated in a *personal* context, belongs to the *space and shape* content area, and classified in the *connection* competency cluster. The results show that the item has a difficulty of 503 score points, which places it at Level 3 proficiency. On the space and shape scale, 51% of all students across the OECD community can perform tasks at least at Level 3.

Country	Students' Choice of Correct Judgments					
	Four (Full)	Three	Two	One	None	No Res.
Australia	68.6	14.1	7.2	6.4	2.4	1.2
Canada	69.6	14.0	7.3	6.3	2.1	0.6
Germany	69.0	13.9	7.3	5.6	2.3	1.9
Japan	83.3	8.9	4.2	2.0	0.9	0.7
OECD Average	63.0	16.0	8.9	7.2	2.7	2.3

Table 2: The percentage of students' response for NUMBER CUBES (An excerpt from National Institute for Educational Policy Research, 2004, p. 108.)

The result shows that Japan's mean score was significantly higher than the OECD average as well as being higher than other participating countries (See Table 2). Also, the pattern of students' choice is slightly different from other countries.

In order to complete the item correctly, we need to interpret the two dimensional object back and forth by “folding” it to make the four planes of the cube mentally as a three-dimensional shape. The item requires the encoding and spatial interpretation of two-dimensional objects. Why did a group of students, once again Japanese students, perform well on this particular item? Does the result suggest that those students have a cultural practice with number cubes, or Origami, inside and outside schools? A further exploration is needed to explain the similarities and differences in students’ responses among participating countries.

There are other insights offered by the recent international studies. The TIMSS2003 collected information about teacher characteristics and about mathematics curricula. The PISA2003 also collected a substantial amount of background information through the student questionnaire and the school questionnaire. These data on contextual variables as well as performance data related to the cognitive test domain give us rich descriptions of the learning environments of the learners.

As was mentioned above, the recent release of the two large-scale international achievement studies provides mathematics education researchers with opportunities for exploring research possibilities in relation to learners and learning. While we need to examine the results from each study carefully, we also need to synthesize the results from different perspectives as a coherent body of description of the reality of the learners.

References

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA’s Trends in International Mathematics and Science Study at the fourth and eight grades. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.*

National Institute for Educational Policy Research (2003). *The report of 2001 national survey of the implementation of the curriculum: Mathematics*. Tokyo: Author.

National Institute for Educational Policy Research (2004). *Knowledge and skills for life II: OECD- Programme for International Student Assessment 2003*. Tokyo: Gyosei

National Institute for Educational Policy Research (2005). *A summary of 2003national survey of the implementation of the curriculum: Mathematics*. Tokyo: Author.

Organisation for Economic Co-operation and Development (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: Author.

Organisation for Economic Co-operation and Development (2004). *Learning for tomorrow’s world: First results from PISA 2003*. Paris: Author.

THE PISA-STUDY: CHALLENGE AND IMPETUS TO RESEARCH IN MATHEMATICS EDUCATION

Michael Neubrand

Dept. of Mathematics, Carl-von-Ossietzky-University, Oldenburg (Germany)

Beyond the results, a large scale study like PISA may also stimulate the area of research in mathematics education. Since an empirical study needs a sound conceptualization of the field - "mathematical literacy" in the case of PISA - mathematics education research and development may benefit from the structures of mathematical achievement defined for PISA. Further research can build upon the work done in PISA.

PISA, the "Programme for International Student Assessment" (OECD, 2001, 2004) came into the public focus mainly for the results and the prospective consequences to be drawn: "All stakeholders – parents, students, those who teach and run education systems as well as the general public – need to be informed on how well their education systems prepare students for life" (OECD, 2004, p 3). However, the PISA study deserves interest also from the point of view of research in mathematics education. This perspective is inherent to PISA: The PISA-report "considers a series of key questions. What is meant by 'mathematical literacy'? In what ways is this different from other ways of thinking about mathematical knowledge and skills? Why is it useful to think of mathematical competencies in this way, and how can the results be interpreted?" (OECD, 2004, p 36)

This paper draws attention to some of the impulses and challenges to mathematics education research coming from the PISA studies. We recognize both, the international study, and the national option in Germany which was based on an extended framework and included additional components.

SYSTEM RELATED DIAGNOSIS OF MATHEMATICAL ACHIEVEMENT

What are the aims of PISA? PISA's main focus is to measure the outcomes of the whole educational systems in the participating countries, and choses, as the most sensible group to investigate, the group of the 15 years olds in the countries. The key question therefore is on the system level: What do we know about the mathematical achievement and its conditions in an *educational system* compared to what one can observe in an international overview?

Apparently, this is not thoroughly in tune to the mainstream of mathematics education research. There are long and ongoing traditions in mathematics education which point to a contrasting aspect: What are an *individual's* thoughts, difficulties, sources, and strategies when learning mathematics? Our common interest is often more on an individual's understanding, or on the misunderstandings in the social communication among the individuals in the classroom. Thus, it does not wonder that international comparisons found and still find critical reactions, going back as far as

2005. In Chick, H. L. & Vincent, J. L. (Eds.). *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 1, pp. 79-82. Melbourne: PME.

Hans Freudenthal's fundamental critique in the beginning of comparative studies in mathematics (Freudenthal, 1975).

Contrasting that tradition, the complementary question towards a *systems' efficiency* in mathematics teaching and learning is not less challenging. One has to define appropriate concepts and instruments to answer the question on a basis which incorporates the knowledge mathematics education research has given us so far. In fact, PISA took that challenge serious in a twofold way: The concept "mathematical literacy" forming the basis for testing mathematics achievement is explicitly bound to the mathematics education tradition (OECD, 2003; Neubrand et al., 2001); and vice versa, the PISA test gave rise to further developments of conceptualizing mathematical achievement (Neubrand, 2004). Thus, PISA provides theoretically based, and empirically working conceptualizations of mathematical achievement, which can be seen as an impetus to mathematics education research.

CONCEPTUALIZING MATHEMATICAL ACHIEVEMENT

Sources of the concept "mathematical literacy"

The specific idea of PISA is that the outcomes of an educational system should be measured by the competencies of the students. The key concept is "literacy". Three roots can be traced back: a tradition of pragmatic education (e.g., Bybee, 1997), Freudenthal's conception that "mathematical concepts, structures and ideas have been invented as tools to organise the phenomena of the physical, social and mental world" (Freudenthal, 1983), and considerations on what mathematics competencies are about (Niss, 2003). From there the PISA-framework developed that PISA aims to test the capability of students "to put their mathematical knowledge to functional use in a multitude of different situations" (OECD, 2003).

Conceptualizing „mathematical literacy" in the international PISA study

The domain "mathematical literacy" was conceptualized and related to the test items (problems) in the international PISA study by three components (Fig. 1).

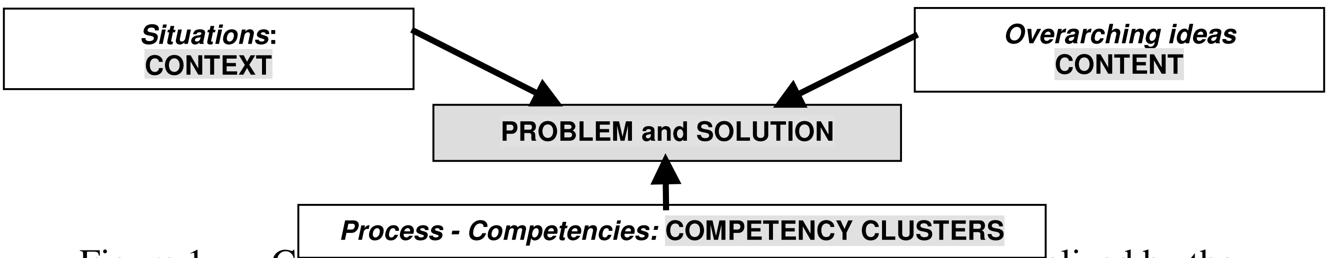


Figure 1. Components of mathematical problems as conceptualized by the international PISA framework (OECD, 2003, p. 30).

It is one of the major impetuses (and challenges) to mathematics education research that (and if) a list of mathematical competencies, accumulated in the Competency Clusters ("Reproduction" - "Connections" - "Reflection"), may hold as "key characteristics" (OECD, 2004, Annex A6) to construct an appropriate instrument to test mathematical achievement. In 2004 PISA reported countries' achievement differentiated by the content-dimension, and it will be a matter of further research to clear how far the competencies itself are present in the countries.

Conceptualizing mathematical achievement in the German national PISA option

Even stronger than PISA-international, the German national option capitalizes that an achievement test like PISA should map mathematics as comprehensively as possible. Therefore, typical ways of thinking and knowing in mathematics should be present in the test items. This model of the test tasks formed the basis (Fig. 2):

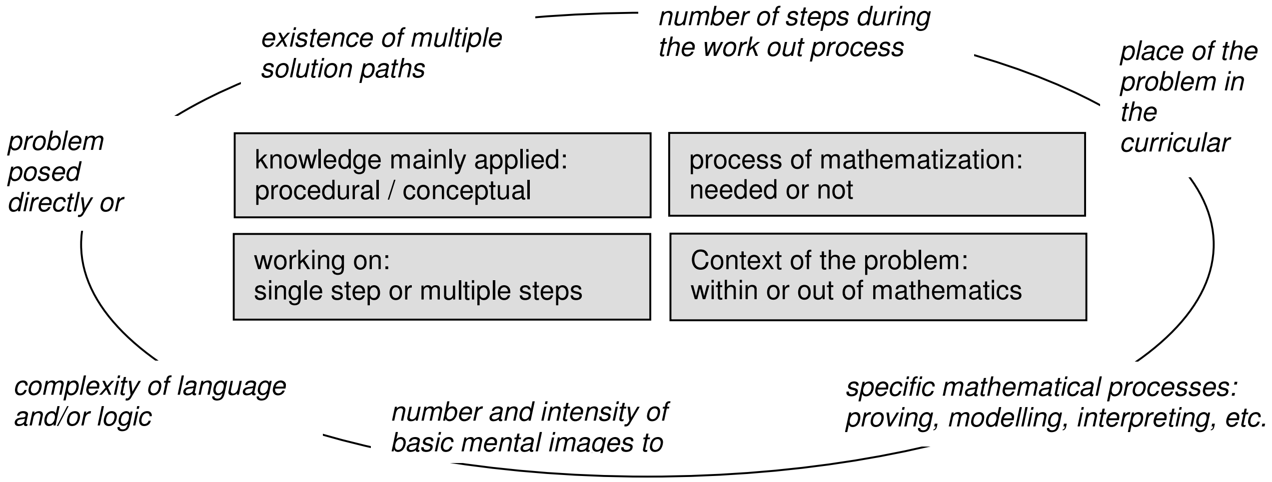


Figure 2. The model of a mathematical problem used in PISA-Germany: The core, and examples of characteristic features (Neubrand, 2004)

With the four basic features (the "core") mathematical achievement can be structured by three "types of mathematical activities" (J. & M. Neubrand in Neubrand, 2004): (i) employing only techniques, (ii) modeling and problem solving activities using mathematical tools and procedures, (iii) modeling and problem solving activities calling for connections and using mathematical conceptions. From the cognitive and the mathematical point of view the three classes realize the full range of mathematical thinking, since one recognizes technical performance, and the essential modes of thinking, i.e., procedural vs. conceptual thinking (Hiebert, 1986).

ANALYTIC RESULTS OF PISA

The defined structures of mathematical achievement express themselves also in the data. But clearly, there remains a lot to do for further research.

International test: Countries show differences in content areas

Not surprisingly, test data show differences among countries in the performance on the defined content areas, the "overarching ideas" (OI). While the students in some countries behave quite uniformly over the content (e.g., Finland, Belgium), in some countries considerable differences appear. For example, Japan shows strengths in the OIs "change and relationships" and "space and shape"; and (relative) weaknesses in "quantity" and "uncertainty". Germany shows weakness in the geometry and stochastics items. Results like these give hints what fields of mathematics should earn greater emphasis in curriculum and teaching. (See OECD, 2004 for details.)

Difficulty of a problem: A question of various features

Analyses done after PISA-2000 in Germany revealed some insight into the processes which make the solution of an item more difficult. However, as said in the beginning, due to the nature of the data, one can get information on mathematical learning and thinking in the whole, and not information of an individual's ways of thinking. Nevertheless, there are interesting results to obtain.

(a) *Not* the same features make a problem difficult in any of the three "types of mathematical activities" (J. & M. Neubrand in Neubrand, 2004). As a consequence, mathematic teaching cannot restrict itself to only a limited scope of mathematics.

(b) There is a competency specific to mathematics, that influences the difficulty of problems, even of those problems which call for modeling processes: the capability to use formalization as a tool (Cohors-Fresenborg & al. in Neubrand, 2004).

(c) Different didactical traditions and ways of teaching lead to different "inner structures" of mathematical achievement, made visible by different performance in the types of mathematical activities (J. & M. Neubrand in Neubrand, 2004).

References

Bybee, R.W. (1997). Towards an understanding of scientific literacy. In W. Gräber & C. Bolte (Ed.), *Scientific literacy. An international symposium* (S. 37-68). Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN).

Freudenthal, H. (1975). Pupils' achievements internationally compared - The IEA. *Educational Studies in Mathematics 6,* 127 - 186.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht: Reidel.

Hiebert J. (Ed.). *Conceptual and procedural knowledge: The case of mathematics*. Hillsdale, NJ: Lawrence Erlbaum.

Neubrand, M. (Ed.) (2004). *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland: Vertiefende Analysen im Rahmen von PISA-2000*. Wiesbaden: VS - Verlag für Sozialwissenschaften.

Neubrand M. et al. (2001). Grundlagen der Ergänzung des internationalen PISA-Mathematik-Tests in der deutschen Zusatzerhebung. *Zentralblatt für Didaktik der Mathematik 33*, 33-45.

Niss, M. (2003). Mathematical competencies and the learning of mathematics: The Danish KOM Project. In A. Gagatsis & S. Papastavridis (Eds.), *3rd Mediterranean conference on mathematical education. Athens, Jan. 2003* (S. 115-124). Athens: Hellenic Math. Society.

OECD - Organisation for Economic Co-operation and Development (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris: OECD.

OECD (2003): *The PISA 2003 Assessment Framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.

OECD (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD.

SOME RESULTS FROM THE PISA 2003 INTERNATIONAL ASSESSMENT OF MATHEMATICS LEARNING: WHAT MAKES ITEMS DIFFICULT FOR STUDENTS?

Carolyn Kieran

Université du Québec à Montréal

Département de Mathématiques

With the announcement of the 2003 PISA results in December 2004, we can now take a closer look at the released items and at how the 15-year-olds of the PISA assessment fared. A brief examination of item difficulty within the "change and relationship" scale suggests that we still know little about what it is that students find difficult in certain mathematical tasks.

MATHEMATICAL LITERACY IN PISA

The PISA concept of mathematical literacy is concerned with "the capacity of students to analyse, reason, and communicate effectively as they pose, solve and interpret mathematical problems in a variety of situations involving quantitative, spatial, probabilistic or other mathematical concepts" (OECD, 2004, p. 37). More precisely, mathematical literacy is defined as "an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen." The objective of the PISA 2003 assessment was "to obtain measures of the extent to which students presented with problems that are mainly set in real-world situations can activate their mathematical knowledge and competencies to solve such problems successfully" (OECD, 2004, p. 57).

HOW MATHEMATICAL LITERACY WAS MEASURED

Students' mathematics knowledge and skills were assessed according to three dimensions: mathematical content, the processes involved, and the situations in which problems are posed. Four content areas were assessed: shape and space, change and relationships, quantity, and uncertainty – roughly corresponding to geometry, algebra, arithmetic, and statistics and probability. The various processes assessed included: thinking and reasoning; argumentation; communication; modeling; problem posing and solving; representation; and using symbolic, formal, and technical language and operations. The competencies involved in these processes were clustered into the reproduction, connections, and reflection clusters. The situations assessed were of four types: personal, educational or occupational, public, and scientific. Assessment items were presented in a variety of formats from multiple choice to open-constructed responses.

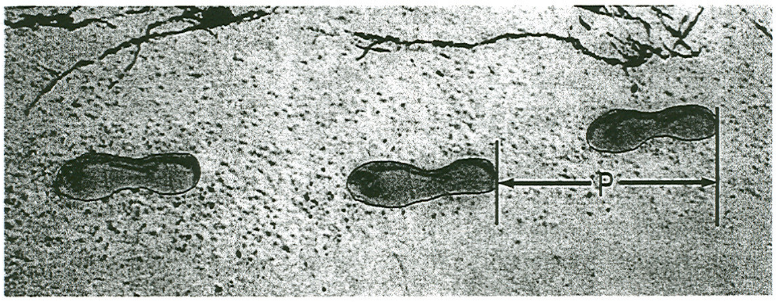
2005. In Chick, H. L. & Vincent, J. L. (Eds.). *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 1, pp. 83-86. Melbourne: PME.

The PISA 2003 mathematics assessment set out to compare levels of student performance in each of the four content areas, with each area forming the basis of a separate scale. Each assessment item was associated with a point score on the scale according to its difficulty and each student was also assigned a point score on the same scale representing his or her estimated ability. Student scores in mathematics were grouped into six proficiency levels, representing groups of tasks of ascending difficulty, with Level 6 as the highest. The mathematics results are reported on four scales relating to the content areas mentioned above. As will be seen, an examination of item-difficulty within these scales reveals some surprises that, in turn, suggest that we, as researchers, may not really know what makes some mathematical tasks more difficult than others for students.

ITEM DIFFICULTIES FOR SAMPLE ITEMS FROM THE CHANGE AND RELATIONSHIP CONTENT AREA: THE WALKING UNIT

The Walking unit (OECD, 2004, p. 64) begins as follows:

WALKING



The picture shows the footprints of a man walking. The pacelength P is the distance between the rear of two consecutive footprints.

For men, the formula, $\frac{n}{P} = 140$, gives an approximate relationship between n and P where:

n = number of steps per minute, and
P = pacelength in metres.

Items 4 and 5 from this unit, along with the respective item difficulties and discussion of the competency demands, are presented in Figure 1. The level of difficulty ascribed to Item 4 is difficult to fathom: 611, which places it at Level 5 proficiency – a level at which only 15 % of OECD area students are considered likely to succeed. Yet, the item requires simply substituting n by 70 in the given formula $n/p = 140$, and then dividing 70 by 140. Its difficulty would seem closer to a Level 2 proficiency item, which according to the OECD report typically involves the "interpretation of a simple text that describes a simple algorithm and the application of that algorithm" (p. 69) – a task that 73% of OECD area students would be likely to solve. While students might attempt to solve the equation $70/p = 140$ by a cross-multiplication technique, they could also think about the task in terms of proportion ($70/p=140/1$, i.e., 70 is to 140 as p is to 1) or arithmetically in terms of division (70 divided by what number yields 140?).

WALKING

QUESTION 5

Bernard knows his pacelength is 0.80 metres. The formula applies to Bernard's walking.

Calculate Bernard's walking speed in metres per minute and in kilometres per hour. Show your working out.

Score 3 (723)
Answers which indicate correctly metres/minute (89.6) and km/hour (5.4). Errors due to rounding are acceptable.

Score 2 (666)
Answers which are incorrect or incomplete because:

- They were not multiplied by 0.80 to convert from steps per minute to metres per minute.
- They correctly showed the speed in metres per minute (89.6 metres per minute) but the conversion to kilometres per hour was incorrect or missing.
- They were based on the correct method (explicitly shown) but with other minor calculation error(s).
- They indicated only 5.4 km/hr, but not 89.6 metres per minute (intermediate calculations not shown).

Score 1 (605)
Answers which give n = 140 x .80 = 112 but no further working out is shown or incorrect working out from this point.

This open-constructed response item is situated in a personal context. The coding guide for this item provides for full credit, and two levels of partial credit. The item is about the relationship between the number of steps per minute and pacelength. It follows that it fits the change and relationships content area. The mathematical routine needed to solve the problem successfully is substitution in a simple formula (algebra), and carrying out a non-routine calculation. To solve the problem, students first calculate the number of steps per minute when the pace-length is given (0.8 m). This requires substitution into and manipulation of the expression: n/0.8 = 140 leading to: n = 140 x 0.8 which is 112 steps per minute. The next question asks for the speed in m/minute which involves converting the number of steps to a distance in metres: 112 x 0.80 = 89.6 metres; so his speed is 89.6 m/minute. The final step is to transform this speed into km/h - a more commonly used unit of speed. This involves relationships among units for conversions which is part of the measurement domain. Solving the problem also requires decoding and interpreting basic symbolic language, and handling expressions containing symbols and formulae. The problem, therefore, is rather a complex one involving formal algebraic expression and performing a sequence of different but connected calculations that need understanding of transforming formulas and units of measures. The lower level partial credit part of this item belongs to the connections competency cluster and with a difficulty of 605 score points it illustrates the top part of Level 4. The higher level of partial credit illustrates the upper part of Level 5, with a difficulty of 666 score points. Students who score the higher level of partial credit are able to go beyond finding the number of steps per minute, making progress towards converting this into the more standard units of speed asked for. However, their responses are either not entirely complete or not fully correct. Full credit for this item illustrates the upper part of Level 6, as it has a difficulty of 723 score points. Students who score full credit are able to complete the conversions and provide a correct answer in both of the requested units.

QUESTION 4

If the formula applies to Heiko's walking and Heiko takes 70 steps per minute, what is Heiko's pacelength? Show your work.

Score 1 (611)
Answers which indicate p = 0.5 m or p = 50 cm or
p = 1/2 (unit not required).

This open-constructed response item is situated in a personal context. It has a difficulty of 611 score points, just 4 points beyond the boundary with Level 4. Everyone has seen his or her own footsteps printed in the sand at some moment in life, most likely without realising what kind of relations exist in the way these patterns are formed, although many students will have an intuitive feeling that if the pace-length increases, the number of steps per minute will decrease, other things equal. To reflect on and realise the embedded mathematics in such daily phenomena is part of acquiring mathematical literacy. The item is about this relationship: number of steps per minute and pacelength. It follows that it fits the change and relationships content area. The mathematical content could be described as belonging clearly to algebra. Students need to solve the problem successfully by substitution in a simple formula and carrying out a routine calculation: if n/p = 140, and n = 70, what is the value of p? The students need to carry out the actual calculation in order to get full credit. The competencies needed involve reproduction of practised knowledge, the performance of routine procedures, application of standard technical skills, manipulation of expressions containing symbols and formulae in standard form, and carrying out computations. Therefore the item belongs to the reproduction competency cluster. The item requires problem solving by making use of a formal algebraic expression. With this combination of competencies, and the real-world setting that students must handle, it illustrates Level 5, at the lower end.

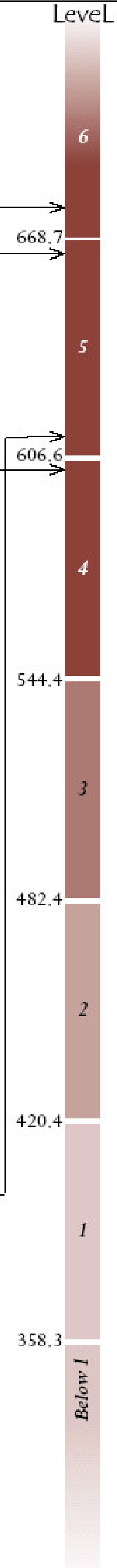


Figure 1. Items 4 and 5 of the Walking unit (OECD, 2004, p. 65)

Curiously, a response earning a partial score of 2 on the seemingly much more difficult Item 5 – at least more difficult from an *a priori* perspective – places it at Level 5 as well, albeit nearer the upper boundary of Level 5. But, it is not clear why a response that is deemed incomplete (and receives a score of 2) because the "112 steps per minute was not multiplied by .80 to convert it into metres per minute" – a conceptual demand that is at the core of Item 5 – is considered superior to the response "n = 140 x .80 = 112," which appropriately receives a partial score of 1. Notwithstanding the argument that could be made for both of these responses" to Item 5 receiving the same score of 1, the main issue concerns the conceptual demands that are inherent in Item 5, but which are lacking in Item 4. Why do students find Item 4 just about as difficult as Item 5?

While some might claim that the procedural demands of Item 4 (with the unknown in the position of denominator) explain to a certain extent why the difficulty level is 611, results from past research studies of equation-solving errors suggest that the difficulty level of this item should not be so high. For example, Carry, Lewis, and Bernard (1980) reported the following success rates for the solving of the given equations among students who covered a range from strong to very weak in algebra skills (e.g., 82%: 9(x+40) = 5(x+40); 76%: 1/3 = 1/x + 1/7; 76% 5/10 = (x-10)/(x+5)). In another study involving classes of 6^{th} to 8^{th} grade students, younger than those tested within PISA, Levin (1999) reported that 30% of the students correctly answered the following question by setting up and solving a proportion using cross multiplication (5/9=2/n): "On a certain map, the scale indicates that 5 cm represents the actual distance of 9 miles. Suppose the distance between two cities on this map measures 2 cm. Explain how you would fine the actual distance between the two cities." The equation was not unlike the one involved in Item 4; moreover, the students had to generate it themselves from the problem situation. One can only conclude that if the PISA results for this item and related symbolic representation items represent a trend with respect to students' abilities to handle rather simple symbolic forms, it is indeed a disturbing one. While Nathan and Koedinger (2000) noted that students find symbolically-presented problems more difficult than story problems and word-equation problems, the PISA results suggest that the discrepancy may be much greater than that reported by these researchers.

References

Carry, L. R., Lewis, C., & Bernard, J. E. (1980). *Psychology of equation solving – an information processing study* (Final Technical Report). Austin: The University of Texas at Austin.

Levin, S. W. (1999). Fractions and division: Research conceptualizations, textbook presentations, and student performances (doctoral dissertation, University of Chicago, 1998). *Dissertation Abstracts International 59*: 1089A.

Nathan, M. J., & Koedinger, K. R. (2000). Teachers' and researchers' beliefs about the development of algebraic reasoning. *Journal for Research in Mathematics Education, 31*, 168-190.

OECD (2004). *Learning for tomorrow's world – First results from PISA 2003*. Paris: OECD.

THE FOUNDATION AND SPECTACLE OF [THE LEANING TOWER OF] PISA

Julian Williams

University of Manchester

I raise questions about the construct and consequential validity of international studies such as PISA, and about PISA itself. I suggest a fault line runs through the construct 'mathematical literacy', but more importantly, through mathematics education generally, distinguishing 'Realistic' mathematics and 'Authentic' mathematics. I then ask questions about the political consequences of PISA in an audit culture in which targets beget processes. The aim to influence policy is identified with perceptible shifts in PISA discourse. As an instrument in the global education market, with its theft of critical theorists' rhetorical resources, is PISA re-invigorating the spectacle of international league tables?

INTRODUCTION

When I was a boy I visited Pisa and was very impressed by the leaning tower. I recall imagining that one could walk up the tower by spiralling up the outside, and was slightly disappointed by the reality. Later I learned that the inclination of the tower was annually increasing, and engineers feared that it would eventually fall over: they planned to strengthen the foundations to stop this, but did not straighten it. The tower has become a global spectacle, even featuring in jokes etc. (what did Big Ben say to the leaning tower of Pisa? I've got the time if you've got the inclination). The tower of Pisa became globally spectacular because of its dodgy foundations, not despite them.

I aim to raise questions about the validity of PISA (capitals now). First, I examine the construct validity of the foundation of PISA, 'mathematical literacy'; second, I address the consequential validity of PISA, its political consequences, as spectacle.

CONSTRUCT VALIDITY: THE FOUNDATION OF 'MATHEMATICAL LITERACY'

A confession: I find some of the items in PISA seductive, especially some of the Problem Solving items. In one the student is asked to diagnose a faulty bicycle pump, in another they are asked to evaluate some information on various drugs and select an appropriate pain-killer for 13 year old George, an asthmatic child with a sprained ankle. At face value, these represent a kind of functional 'literacy'. Turning to the mathematical literacy item used to explain the notion of mathematical modelling and mathematisation, one finds the park problem: where should a street-light be placed to illuminate a park? The park is mathematised as a triangle, the area lit is a circle, and the solution is the triangle's circumcentre (as long as the park is not obtuse-angled, explains PISA, 2003, p26).

2005. In Chick, H. L. & Vincent, J. L. (Eds.). *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 1, pp. 87-90. Melbourne: PME.

I may be obtuse, but … our parks in English towns are usually locked at night, not lit. Perhaps they mean a car park? But … how many triangular car parks have you seen? I looked around and noticed that the lights were often on the perimeter of the park, which is in turn usually made up of rectangular blocks. For obvious reasons one might expect car parks to be rectangular, especially in modern countries where road systems are grid based. Perhaps one would find them in towns where road networks crystallised on the basis of clusters of medieval villages, like Chester or York? Both these towns are a long way from Manchester, so this prompted me to email my co-presenter from Japan and… he found one! (But where was the lighting?)…

Does the validity of Euclid really lie in such considerations? How has this come to be? I fantasise: Euclid, on a trip to visit the leaning tower, finds a triangular car park and noticing the light at the midpoint of one side… "Eureka: the circumcentre of a right-angled car park lies at the mid-point of the hypotenuse."

But Realistic Mathematics Education (RME) does not require that mathematics be authentic in this 'real' sense: only that the situation is realistic for the entry of the student into a world that begs to be mathematised. The validity test for RME then is (i) mathematical, rather than 'real' functionality, and (ii) empirical (i.e., do the students experience the problem in an intuitive way). Many of the PISA items appear to have this quality, at least to some degree.

I suggest that Realistic mathematics is primarily embedded in a scholastic, pedagogical activity system and is essentially embedded in the students' imaginary, experiential world: the object of activity is, in the end, to learn mathematics. On the other hand, I suggest Authentic mathematics is used as an instrument within an Activity System whose object is not essentially to learn mathematics, but to achieve some 'real' objective in a world outside mathematics. To become Authentically functional is to break out of the scholastic straitjacket and requires what Engestrom (e.g., Engestrom, 1987) called 'expansive' activity: at the very least, the class that 'plans a party' has to really have the party.

I prefer to think of this distinction as a fault line deep underneath the surface of the concept of 'mathematical literacy', rather than a dichotomy as such. Does this line undercut the mathematics education literature too?

And where is PISA? I'd say some of the best tasks are Realistic, but never quite Authentic (you would hope George's 15 year old literate elder sibling would think to ask a good pharmacist before deciding which painkiller to buy his asthmatic younger brother, wouldn't you? Sorry, 'code 0: no credit'). Could they be?

DISCOURSE AND SPECTACLE OF PISA: POLITICAL CONSEQUENCES

PISA has a political aim, that is, it seeks to influence policy. Thus on the one side, we have mathematics-literacy tasks, and the identification of learning outcomes for students. But on the other, we have summative statistics that 'count' for policy. This entails an interesting discursive shift. Initially, PISA (e.g., 2005) suggest that correlations display ‘associations’ that cannot be assumed to be ‘causal’, but later these associations become ‘influences’ that policy makers might find ‘interesting’. What is the difference for policy, i.e. what is the political difference between an influence and a cause? I see from the dictionary (OED) that an influence is in its original usage an astrological one, and later became political: it is essentially the exertion of an action whose mechanism is ‘unseen’ except in its effects.

This is significant because it determines to some extent the ‘consequences’ of PISA. How can policy makers be expected to read PISA’s results on the influence of SES or softer variables such as ‘school climate’ on learning outcomes? We see from the PISA-2000 study, for instance PISA (2005), that school climate explains significant variation in outcomes, but not that school climate is a possible ‘associate’ of high learning outcomes, and in Gill et al. (2002), associations with school background become ‘attributable’ to school background (p xvi).

Michael Power, who calls himself a professor of critical accountancy, has described the discourse of performativity in our audit culture (i.e., that of managing targets, league tables, performance-related reviews, etc.) as a Foucaultian discourse of (mis-) trust (Power, 1999). He and others have pointed to the way measurement constructs become targets and begin to dominate processes: thus as I write Prime Minister Tony Blair is felled by an angry electorate in debate on TV. He is accused of being responsible for the fact that in some doctors’ surgeries patients are not allowed to book an appointment to see their doctor more than 2 days ahead. Why? Because the government had introduced a performance target for the percentage of patients that have to wait more than 2 days. In vain he protests that this was not his intention! How will PISA measures be used, and what will be their unintended consequences?

Stronach (1999) in ‘Shouting theatre in a crowded fire’ construes the international tests and league table performance as a global spectacle, with ‘pupil warriors’ doing their sums for Britain. There’s England in the Premier league, 3 up on old rivals Germany, there’s a cluster of Confucian Pacific rim teams in the lead, but here comes Finland from nowhere suddenly challenging them. Is it social democracy or Nokia that ensures the team’s strength?

The association between PISA/TIMSS league tables and football competitions, the Olympics, horse races etc. is too strong to be denied, and ‘England’ in the tables becomes metonymically the nation and its education system per se, competing in the game with the rest of the world. One forgets that in fact the order of the names in the table are mostly not statistically significant, of course. What else is a table of scores actually for except to emphasise the ordinal at the expense of the complexity of the underlying data/reality? (That is intended to be a mathematically literate observation, if you like.)

The tabloid/redtop press are masters of this spectacle, but we all become implicated: government funding for research (at least in the UK) is increasingly predicated on ‘making a difference’ to learning outcomes in practice, and hence fulfilling political demands to become ‘world class’. But how can world class be judged, except by international competition and league tables, and hence comparative measurement?

With what consequence? Is there no going back? Has the spectacle seduced our rationality? Pisa will always be the place with the leaning tower. While PISA challenges TIMSS by engaging with some ‘literacy’ rhetoric drawn from critical theory, the source of much that seems seductive in it, one reading of this move might be, as Gee et al. (1996) and others have suggested with ‘fast capitalism’, that the system steals critical theorists’ rhetorical resources and emerges all the stronger for it.

So, where next? Could an expanded Authentic mathematics assessment emerge to confront the Realistic PISA, and in whose interest might that be?

References

(see www.education.man.ac.uk/lta/pme/PISA)

Engestrom, Y. (1987). *Learning by expanding: an activity-theoretical approach to developmental research*. Helsinki, Orienta-Konsultit.

Gee, J., Hull, G. and Lankshear, C. (1996). *The new work order: behind the language of the new capitalism*. Westview Press, Boulder, CO.

Gill, B., Dunn, M. & Goddard, E. (2002). *Student achievement in England*. London: HMSO.

PISA (2003). *The PISA 2003 assessment framework*. Paris: OECD.

PISA (2004). *Learning for tomorrow’s world*. Paris: OECD.

PISA (2005). *School factors related to quality and equity*. Paris: OECD.

Power, M. (1999). *The audit explosion: rituals of verification*. Oxford: OUP.

Stronach, I. (1999). *Shouting theatre in a crowded fire*. Evaluation, 5(2), 173-193.

Afterword

Is the metaphoric association of Pisa and PISA – their foundations and their glorious spectacles – valid? If the consequence is that one is inclined to believe that there is a fault underlying ‘mathematical literacy’, I suggest yes. If one is led to think that this fault is implicated in the faux-spectacle of PISA, perhaps: the argument is that the act of global assessment becomes false *by virtue* of its becoming a political spectacle.

[Acknowledgements: to Google.com for suggesting the Pisa=PISA metaphor, and Ian Stronach for the introduction to this notion of spectacles.]