TIMSS VERSUS PISA: THE CASE OF PURE AND APPLIED MATHEMATICS

Liv Sissel Grønmo, University of Oslo, Norway
Rolf Vegar Olsen, University of Oslo, Norway

Abstract

Framework, aims and population differs in TIMSS (8th grade study) and PISA. We see this as an opportunity to get more knowledge and insight in the educational system in different countries than one study alone could offer. Several countries participated in both studies in 2003. By comparing the countries' ranks in the two studies it is evident that a group of countries, particularly some Nordic and English-speaking countries, perform relatively better in PISA. On the other hand, the East European countries perform relatively much better in TIMSS. An analysis of the mathematical coverage in the two studies has been done in order to understand these shifts in rank. The findings of our analyses are: (a) the assessment frameworks are formulated from largely different perspectives on mathematics. While PISA describes in detail the contexts and phenomena where mathematical competence may be of importance, TIMSS gives a very fine-grained definition of some important aspects of mathematics from within the discipline itself. (b) The items in PISA emphasises the use of numerical information in the form of tables and graphs taken from real world contexts, while items in TIMSS give much more attention to pure mathematics, including formal aspects of algebra and geometry. We also present country characteristic profiles across major categories in TIMSS and PISA for five selected countries. Based on these results, we discuss the relation between pure and applied mathematics in school, and conclude that to do well in applied mathematics, it is necessary with a basis in elementary knowledge and skills in pure mathematic. For some countries like Norway, it seems to be most problematic that students lack elementary knowledge and skills in a topic as Number.

INTRODUCTION

At the outset PISA and TIMSS are regarded as very similar types of study. They are large-scale surveys with a very similar methodological basis, e.g. they

- are sample-based studies of clearly defined populations;
- apply the same type of instruments (e.g. student questionnaire and cognitive booklets);
- process the data with similar psychometrical methods;
- are governed by a consensus-driven process from initial ideas to final instruments;
- enforce rigorous quality control, e.g. of translation or adaptation of the test material;
- have cyclic designs with a focus on measuring trends.

Furthermore, both studies include measurements of highly related constructs: e.g. mathematical and scientific competency, and student and school background characteristics and attitudes. However, we have here used terms like "similar" and "same" and not equal, identical or equivalent. There are important differences between the studies, and in the following some of these will be highlighted instead.

The problem of this apparent similarity between the studies was particularly felt in December 2004. Both TIMSS and PISA published the results from their studies conducted in 2003 (Martin et al., 2004; Mullis et al., 2004; OECD, 2004) within a few days this month. Both studies were followed up with intensity in the media for a period of time, and regularly media referred to both through statements like: "An international study of mathematics achievement shows that students in our country…". Moreover, both studies referred to students at the same educational level, (lower) secondary school, in most countries. It is important to communicate that the key terms 'international study', 'mathematics' and 'students' are not synonymous when referring to PISA or TIMSS respectively. This paper will compare and contrast these studies, and we will particularly discuss how the keyword "mathematics" has been operationalised differently in the two studies. Although the case of Norway will be central in the paper (Grønmo et al., 2004; Kjærnsli et al., 2004), the overall relevance and purpose of it go beyond a national context.

Relevance of the study

There are several reasons why such a comparison is relevant to make:

- It will emphasize how the monitoring of educational systems[1] by international comparative studies may be done from different perspectives, leading to different results and possibly different interpretations.
- The two studies represent two partly overlapping and partly diverging views on mathematics education. Furthermore, by comparing the two studies one might also unravel or discover more tacit, or rarely spoken of, aspects of the two studies' design.
- Several countries participate, have participated, or will be participating in both studies. Of the 32 educational systems that took part in PISA 2000, 28 also participated in either TIMSS 1995 or TIMSS 1999, or both. This number is even higher for countries participating in PISA 2003. In 2003 22 educational systems took part in both studies.
- For researchers in mathematics education the fact that many countries have participated in both studies may yield new opportunities for studying international differences and similarities in mathematics education.

The importance of analyzing the relationships between these studies has also been recognized by others. OECD has commissioned a thematic report to compare PISA and TIMSS (not published yet), and in the US context the National Center for Education Statistics also has ongoing efforts to compare TIMSS and PISA, and furthermore to relate the studies to the NAEP studies (National Center for Education Statistics (NCES), 2005; Neidorf et al., forthcoming; Nohara, 2001).

In this paper we will present the results of comparisons of frameworks and items in TIMSS (8th grade) and PISA. Furthermore, we present the achievement profiles across different content domains for five countries selected to represent the international diversity in achievement profiles. These results give us a background for discussing differences in mathematics education in different countries, beyond what data from one of the studies could offer. A main topic in the discussion will be the perceived tension between elementary mathematical knowledge and skills versus problem solving in mathematics, including a discussion of the relationship between pure and applied mathematics.

What constitute school mathematics?

There has been an ongoing discussion about what should constitute mathematics in school after the Second World War, to a large extent based on the great effort and use of resources to develop education for all citizens in Western societies (Ernest, 1991; Skovsmose, 1994). Central to this discussion has been the relationship between pure and applied mathematics in the school curriculum. There have been several "back to the basics" movements, particularly in the US context, and the discussion has been so heated that the label "math wars" has been frequently used (Schoenfeld, 2004). We will argue that the different operationalizations of mathematics in TIMSS and PISA to a large degree reflects this discussion about the relation between on the one hand mastery of basic facts and procedures in mathematics, and on the other, to be able to relate mathematics to real world phenomena.

Figure 1 is a model for how a problem situation in the real world is transformed into a mathematical problem and subsequently how a solution is found and validated against the original real context where the problem originated. The model is taken from a very influential US policy document on standards in mathematics (NCTM, 1989). This model is also adopted by PISA in a slightly different version (OECD, 2003, p. 38). The right-hand side of Figure 1 represents the mathematical world, an

[1] Mostly the participating educational systems are countries. However, there are also some autonomous states or regions within some countries included. In the following we will use the term country to include all participating systems, both formal national states and autonomous regions with independent participation status in the projects.

abstract world with well defined symbols and rules. The left-hand side represents the real concrete world, what we may call daily life. Working with pure mathematics, as numbers or algebra out of any context, means working only on the right-hand side of Figure 1. In applied mathematics, the starting point is supposed to be a problem from the real world, which first has to be simplified, and then mathematized into a model representing the problem. In most cases school mathematics will rarely start with a real problem. In most cases the problem presented to the students has already been simplified.

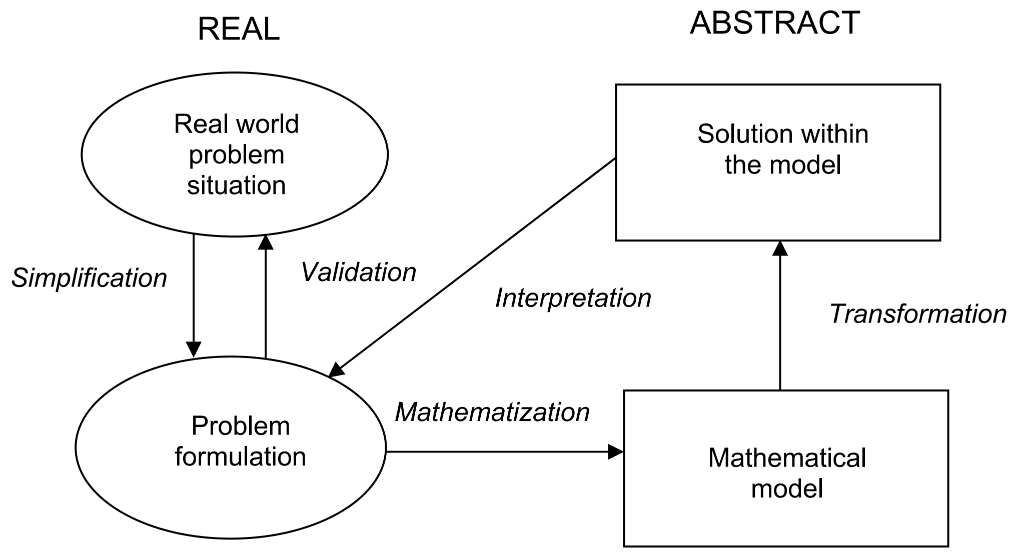


Figure 1: *The mathematization cycle. Copied from the Standards (NCTM, 1989).*

It is obvious from them mathematization cycle in Figure 1 that a premise for any type of applied mathematics is the assumption that the students have the necessary knowledge in pure mathematics to be able to find a correct mathematical solution. The students have to be able to orientate themselves in a pure mathematical world. This illustrate that in any case, applied mathematics is more complex than pure mathematics, taking as a premise that it is the same mathematics that is involved in the two cases. With an increasing focus on applied mathematics in compulsory school in some countries, it seems that to some extent the importance of understanding that mathematics as such is an exact and well defined science, and that the importance of being able to orientate in the world of pure mathematics has been neglected to some extent. This has been criticized by some researchers. Gardiner (2004) points out that even if mathematical literacy, the ability to use mathematics to solve daily life problems is a main goal for school mathematics, this should not be seen as an alternative to basic knowledge and skills in pure mathematics.

The different ways that mathematics is operationalised in TIMSS and PISA reflect different opinions of what school mathematics is or should be. PISA is testing students in the type of applied mathematics they may need in their daily life as citizens in a modern society, it is an aim for PISA to embed all items in a context as close to a real life problem as possible. Most of the items in TIMSS however, are either pure mathematical items with no context, or items with the simple and artificial contexts that are almost a signature of traditional mathematical problems in school. TIMSS will therefore to a great extent give information about students knowledge in pure mathematics, while PISA will display students knowledge in applied mathematics.

The discussion of the results of the analysis of data from TIMSS and PISA will be related to this introduction about the relationship between pure and applied mathematics. It differs to a large extent how this relationship between pure and applied mathematics seem to be represented in school curriculum in different countries. We will not discuss this based on any analysis of the formal mathematical curriculum in different countries, what may be seen as the intended curriculum. The TIMSS framework (Mullis et al. 2003) is based on a model containing three levels of the curriculum, the intended, the implemented and the attained curriculum. Our discussion will be based on analyses of the attained curriculum, what students have achieved of mathematical knowledge, which we will take as indicators for what has been in focused in instruction in school, what we may call the implemented curriculum.

Comparison of general features in PISA and TIMSS

Although both PISA and TIMSS have many overlapping features in their design and administration, there are some important differences which have to be addressed before any meaningful comparison between the two studies definitions and operationalisations of students' achievement in mathematics may be done.

First of all, it is important to note that the two studies are based on two largely different visions for how educational outcomes may be monitored. TIMSS is designed to specifically target what is perceived to be the best common core of mathematics curriculum across the world. In other words, TIMSS aims at measuring educational systems according to the stated aims for those systems. Large efforts are therefore made, based on reviews of the items and frameworks in each country, to formulate a framework representing this "best common core". The principle followed in PISA is rather different. Instead of basing the framework on the curriculum in the participating countries, PISA have challenged leading scholars in mathematics education to define mathematics according to what students will need in their present and future life. The label used is therefore *mathematical literacy* in order to distance the concept from what most people think of when they hear the term *mathematics.* For most people mathematics refers to either a school subject and/or an academic discipline. The term "mathematical literacy" is chosen in PISA to emphasise how the study aims at measuring how well students will be able to cope with situations where mathematical skills and knowledge are of importance.

However, in the public domain the term mathematics is convenient to use, and thus, the distinction between the two studies is often not present when results are discussed in newspapers articles, among policy makers or even among academic scholars. While it is reasonable to say that TIMSS measures how well a system have implemented the mathematics curriculum, it is more reasonable to claim that PISA may be perceived more of as measure of how well the curriculum serves the general students' needs in their current and future life as citizens. The policy implications that may be drawn from the two studies are therefore in principle rather different. While TIMSS may be used to study the implementation of the core curriculum, PISA may be used to evaluate whether the curriculum is well-suited in order to foster citizens equipped with the mathematical competencies that they will most likely need.

The visible effects of these two different perspectives is that mathematics as defined, operationalised and reported in TIMSS is much more related to understanding of fundamental concepts and procedures in mathematics. This is for instance reflected in the very fine-grained and detailed framework, particularly for the content dimension. TIMSS has five content domains in mathematics, e.g. "Number". Each content domain is partitioned by a set of more specific objectives, e.g. "Integers" (one of four objectives within "Numbers"). Finally, each objective is operationalised in more specific statements, e.g. "Compare and order integers" (one of five statements defining the objective "Integers") (Mullis et al., 2001, pp. 12-13). Thus, the defining statements of mathematics achievement describe very specific knowledge and skills. PISA defines mathematics with less rigour on the content dimension. Instead of defining different content domains within mathematics, PISA describes four classes of phenomena where mathematical knowledge and reasoning is of great importance. These are labelled "Overarching Ideas". These four phenomenological aspects are not subdivided any further. They are instead defined by a coherent introductory text giving reference to this class of phenomena generally, and in addition the framework refer to a number of specific situations and many specific examples of items that could be used to assess each of the categories (OECD, 2003). In conclusion it might be stated that:

- the "operational distance" from framework to the specific items is less in TIMSS than in PISA.
- the content subdomains in TIMSS are conceptually more isolated than the overarching ideas in PISA.
- the TIMSS content subdomains clearly demarcate a line between included and excluded content, while the PISA overarching ideas do not draw such a line.

- the TIMSS framework may be captured by the slogan "What school mathematics is", while the PISA framework may be considered as "What school mathematics should be".

The PISA framework gives more attention to the process dimension of using mathematics. The specific focus of this article is on the content dimensions of the two studies (what type of mathematics is involved in the studies?). We will therefore not discuss in details the differences between the cognitive dimensions in the two studies. In order to have a complete comparison of the two studies it is, however, also important to study the differences also along this process or cognitive dimension. Based on our first-hand experience with reviewing, marking and analysing mathematics items in the two studies, it is evident that the cognitive complexity of the PISA items is higher than that of TIMSS. This is also confirmed by a panel reviewing the items on behalf of the US National Center for Education Statistics (NCES), 2005). A majority of the PISA items would be classified as "Reasoning" in the TIMSS 2003 framework.

We have reason to believe that the cognitive dimension do not reflect variation between countries' profiles very clearly. This seems to be supported by the international study on scaling the TIMSS mathematics items across the sub-domains of the cognitive domain (Mullis et al., 2005). The variation within countries across the sub-domains of the cognitive domain is overall lower than across the sub-domains of the content domain. The reasons for this might be several, but we believe that one of the main reasons is that the process or cognitive dimension is much harder to subdivide into distinctly different subclasses. Persons confronted with the task to classify a specific item along both the content and the cognitive domain subcategories would most likely do so with a higher degree of agreement within the content domain, while they would tend to disagree more often in their classifications of the cognitive domain. This is likely due to the fact that an item addresses different cognitive domains for different students. An item for instance testing factual knowledge is designed to separate students who know this fact from those who does not. Nevertheless, students who do not know this fact might be able to reach the correct solution by instead using reasoning skills. Another obstacle in creating meaningful scales across the subdomains of the cognitive dimension is related to the inevitable conceptual hierarchy. Even if an item targets higher order cognitive skills, they will inevitably also to a large degree include important elements of lower order cognitive skills. So, there is a degree of hierarchy in the cognitive domain which makes the task of assessing items as belonging to one of several mutually exclusive sub-categories very difficult. In our account of countries' profiles across different domains in mathematics we will therefore mainly focus on the profiles across the content domains.

In addition to measuring different constructs, there are a number of other important differences between the two studies which may be briefly mentioned. For a fuller account of these and other differences see Olsen (2005):

- PISA targets students who in general are about two years older than students in the TIMSS 8th grade study. Furthermore, TIMSS has a grade based population, while PISA has an age based population. Thus, in TIMSS ages of the students may vary within and across countries, while in PISA grades may vary within and across the countries.
- TIMSS sample classes within schools, while PISA samples students within schools, and thus, the data reflect different hierarchical structures.
- Through the questionnaires the studies collect data on background variables used to supplement, augment and explain the achievement measures. Since TIMSS aims at measuring the implementation of the curriculum, the questionnaire emphasise more heavily than PISA issues relating directly to what happens in the teaching of the subject matter. PISA, on the other hand, puts more efforts into measuring students' social backgrounds.
- The compositions of participating countries are different in the two studies. The range of countries from different developmental levels is wider in TIMSS than in PISA which makes the international comparative backgrounds of the two studies very different.

METHOD

The final aim of this study is to describe and account for some countries' apparent non-consistent performance in the TIMSS 8th grade mathematics assessment and the PISA assessment of mathematical literacy in 2003. Some countries seem to perform relatively better in one of the studies.

It is not possible to combine the measures of achievement in the two studies at the student or school level in any meaningful way. The two measures can not be linked in any formal sense. What we instead can do is to link the two studies at country or system level. In the analysis in this paper we have included only those countries/educational systems with a reported score in both studies. This means that 22 countries and regions are included in the analysis. Although the ranks in the studies are only ordinal measures of achievement, they comprise the only "scale" which is common to both studies. Accordingly, we use the rank orders from 1 to 22 for these countries on both studies independently. The difference in rank is used as a rough indicator of relative success in one of the studies as compared to the other.

Having done that, our aim is to identify possible sources internal to the studies themselves for the shifts. With "internal sources" we refer to characteristic features of the items used to measure mathematics achievement in the two studies. More specifically we have mapped all PISA mathematics literacy items into the mathematics content domain of the TIMSS framework for the 8^{th} grade population. It was decided to conduct two independent classifications by both authors. Disagreements were then resolved in order to reach one final classification.

Variable name	Description
Item format	Classification of items into the three main formats: Selected Response (SR), Short Constructed Response (SCR) or Extended Selected Response. The difference between the two constructed response formats is that the former are items where the response is either one word or one number, while the latter requires a description, an explanation, an argument etc.
Algebraic expressions	Classifies the items that include the use of explicit algebraic expressions.
Calculations	Classifies the items that require exact calculations.
Graphics	Classifies the items that include the use of graphical representations of quantities.
Tables	Classifies the items that include the use of representations of quantities in tables.

Table 1: *Description of some broad (external) descriptors to classify items*

In addition to classifying the PISA items according to the TIMSS content domains, a set of study-independent item descriptors were used as external criteria for comparison of the two studies. All the mathematics items in both studies were classified according to a set of five item descriptors, presented in Table 1 below. Olsen & Grønmo (2006) used these (and some other) indicators in a cluster analysis of countries relative achievement profiles in mathematics in PISA. The indicators could successfully be used to develop a meaningful description of the relative strengths and weaknesses in the different country clusters' achievements in mathematics based on PISA. Since the present study's final aim is to evaluate countries relative success in PISA vs. TIMSS, these descriptors may give us valuable clues to understand why some countries seem to perform relatively better in one of the two studies.

These descriptions of the possible differences between the studies will be discussed in light of the shifts in ranks, and in light of some selected countries relative achievement profiles across different types of content in the two studies.

Results

Country ranks in the two studies

Although the ranks do not allow for the studies of the distances between countries achievements, it is the only viable 'scale' to use for this comparison since there is no formal link between the scales[2].

[2] In order to judge the appropriateness of using this rank, we have calculated the leaps in the international scores as one moves one step in the rank order. On average a shift of one position in the ranks for each of the two studies represent 6 points on both of the the internationally centered scale (both scales with an international mean of 500 and standard deviation of 100) with some large deviations from this mean for the differences between

Table 2 presents the ranks (and the scores) for each of the educational systems participating both in PISA and TIMSS. The difference in rank, PISA rank minus TIMSS rank, is shown in the rightmost column, and the table is sorted from high to low differences. In other words, the countries in the upper part of the table perform relatively better in PISA than in TIMSS, and the countries in lower part of the table perform better in TIMSS than in PISA.

	PISA Score	PISA RANK	TIMSS Score	TIMSS RANK	DIFF RANK
Scotland	524	8	498	15	7
New Zealand	523	10	494	16	6
Norway	495	14	461	20	6
Basque Country, Spain	502	12	487	17	5
Belgium (Flemish)	553	1	537	5	4
Australia	524	8	505	12	4
Sweden	509	11	499	14	3
Netherlands	538	4	536	6	2
Ontario Province, Canada	530	7	521	8	1
Hong Kong, SAR	550	2	586	2	0
Indonesia	360	21	411	21	0
Tunisia	359	22	410	22	0
Quebec Province, Canada	537	5	543	4	-1
Italy	466	19	484	18	-1
Serbia	437	20	477	19	-1
Korea, Rep. of	542	3	589	1	-2
Japan	534	6	570	3	-3
USA	483	16	504	13	-3
Slovak Republic	498	13	508	9	-4
Latvia	483	16	508	9	-7
Hungary	490	15	529	7	-8
Russian Federation	468	18	508	9	-9

Table 2: *Ranks from 1 to 22 for the countries participating in both TIMSS and PISA. The column labelled with "Diff. rank" gives the difference between these ranks, and positive numbers corresponds to higher rank in PISA.*

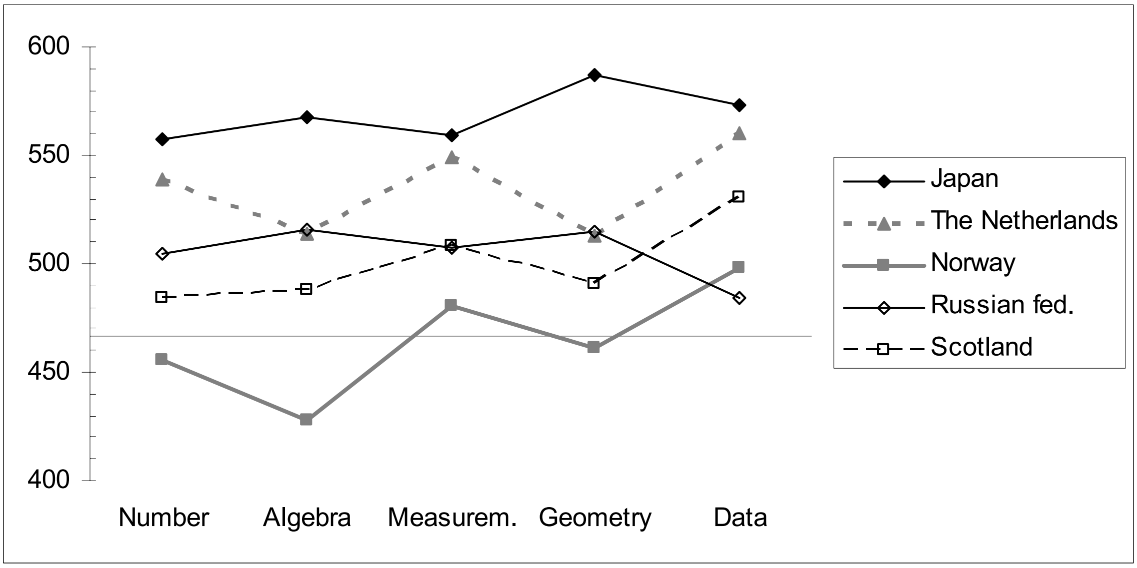
The overall tendency in the country rankings in the two studies is a strong correspondence ($r = 0.76$). The high achieving countries are more or less the same in both studies. However, there is also a systematic tendency in how countries' ranks differ between the two studies. Of course the number of countries is small, and to generalize this pattern to include even more countries is not entirely possible. Still, it is interesting to note that the tendency is that some of the English-speaking countries, the two Scandinavian countries and the Dutch/Flemish speaking educational systems, in addition to Basque in Spain, are ranked relatively higher in PISA. What is even more striking is the tendency for East-European countries to perform relatively stronger in TIMSS.

Country profiles across content

Figures 2 and 3 show the country characteristic profiles across the major content categories in TIMSS (8th grade study) and PISA for five selected countries. These five countries are selected because each of them represents a group of countries with similar achievement profiles in mathematics as documented in the above mentioned cluster analyses. Based on these analyses it is reasonable to highlight some particularly stable groups of countries representing different profiles in mathematics education in school;. We may also talk about a, although this group is not equally well documented in

some of the very low and high ranks respectively. The jump between rank 20 and 21 for the two studies, that is, the jump down to Indonesia, represented an extreme value since the two low performing countries Indonesia and Tunisia are extreme cases on this list.

these studies. In Figures 2 and 3 each of these five groups are represented by the specific country



mentioned in parenthesis above.

Figure 2: *Achievement in major content areas in TIMSS – 8th grade. The first axis crosses the second axis at the overall international mean value of 467.*

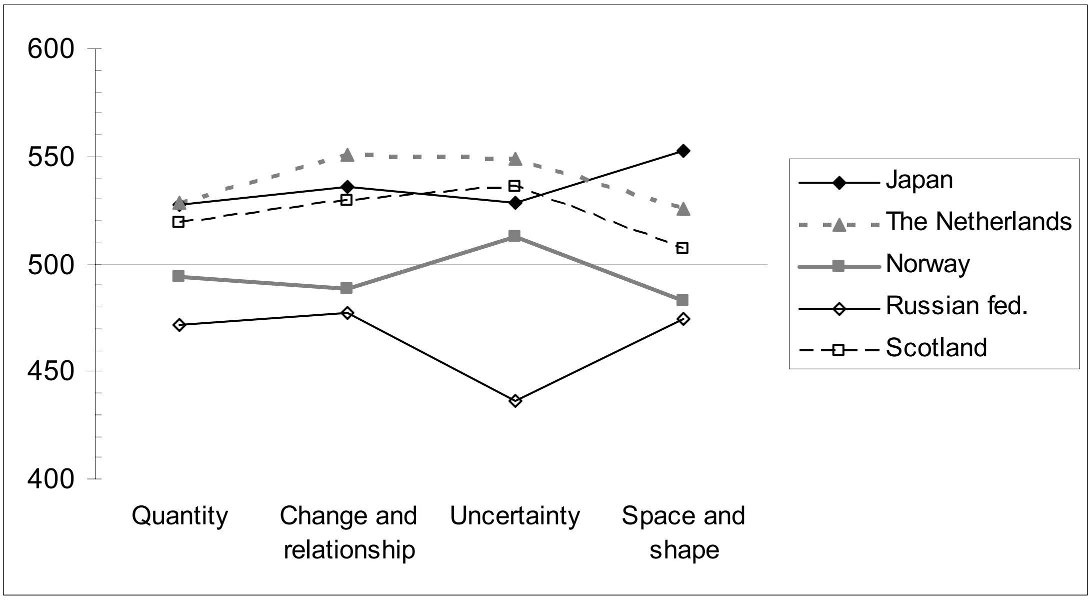


Figure 3: *Achievement in major content areas in PISA. The first axis crosses the second axis at the overall international mean value of 500.*

In short three criteria were used in the selection of these five countries:

- The selected countries represent an international variation in domain specific achievement patterns. The five selected countries each represent one of five well established clusters of countries with distinctly different achievement profiles across the mathematics items in both PISA and TIMSS (Grønmo et al., 2004; Olsen, 2006; Zabulionis, 2001): an English speaking group (Scotland), an East-European group (Russia), an East-Asian group (Japan), a Nordic group closely related to the English speaking group (Norway), and a continental European group (the Netherlands)[3]
- The selected participated in both TIMSS 2003 and PISA 2003.

[3] A specific problem with this group regarding the analyses in this paper, is that most of the predominantly German speaking countries in the group did not participate in TIMSS 2003.

– One possible confounding variable when interpreting data from TIMSS are the age of the participating students. The five selected countries have comparable average age in the TIMSS 8th grade study.

There are many notable features when comparing the overall tendencies in Figure 2 and 3. Firstly, the between country variation is larger in TIMSS than in PISA. The data from TIMSS included in Figure 2 spans the achievement scales from about 420 to 590, while the corresponding PISA data in Figure 3 ranges from 440 to 550. Secondly, the variation across the content domains for each country is larger in TIMSS than the variation between the overarching ideas in PISA. For the TIMSS data in Figure 2 the range for the achievement scores for each country averages to 45 points, while in Figure 3 the same measure is 30 points. Thirdly, the two figures illustrate the overall patterns in the shift of ranks between TIMSS and PISA as presented in Table 2 above: Scottish 15 year-olds perform relatively much better in the PISA assessment than in the TIMSS assessment, while Russia has the opposite shift, scoring relatively much better in TIMSS. A fourth overall observation is the consistent high performance of both Japan and the Netherlands on both assessments across all the mathematical content. For a country like Norway it is very interesting particularly to study in more detail the Dutch mathematics performance. The relative profiles of both countries are almost identical across the content domains in TIMSS (Figure 2) and there are also many similarities in the profiles across the overarching ideas in PISA (Figure 3).

Furthermore, we may note some particular country specific elements in the profiles. Japan, besides scoring very high on all subscales of content in both studies, performs extraordinarily well within Geometry and Space & Shape. For The Netherlands we can see that they perform very well in all domains, but relatively they perform worse in Algebra and Geometry, the two content domains including most formal and abstract mathematics. Norway performs overall very low as compared to the other countries in TIMSS. However, the performance is relatively much better as compared to these countries in PISA, but still towards the lower end of the spectrum. Particularly, the Norwegian performance is characterised by being very weak in Algebra, and relatively much stronger in Data and Measurement in TIMSS and Uncertainty in PISA. The Russian profile has a characteristic minimum for items within Data in TIMSS and Uncertainty in PISA, and the profile is relatively higher for Algebra and Geometry in TIMSS.

Comparing the items in the two studies

It is not entirely easy to compare the instruments used in TIMSS and PISA. First of all – what should be the comparative criteria? As already discussed in the introduction, the PISA items were developed according to content categories reflecting what is labelled as overarching ideas. As briefly presented above, these labels and the whole framework of PISA suggest that the concept of mathematical literacy should be operationalized through more realistically contextualized mathematical problem solving tasks. In general this means that many problems include different types of mathematics. We could say that while TIMSS items have “high fidelities”, the PISA items have broad “bandwidths” in terms of content coverage. Several PISA items could therefore be classified according to more than one TIMSS content domain. One hypothetical PISA item could for instance be a realistic problem relating to the overarching idea “Space and Shape”. Items in this category may typically involve geometrical objects with measurements, and the task may require students to perform calculations with decimal numbers. In our classification of this item into the TIMSS framework, we would therefore have to define it as predominantly either a “Geometry”, a “Measurement” or as a “Number” problem. For most items this was easy, while for some items this was extremely difficult, and in a few cases even impossible.

Table 3 summarises the results of this classification work. This table reports the final classifications after agreement between the two authors. Furthermore, the table summarises the classifications according to the external criteria referred to in Table 1. The table supports the previous brief comparison of the frameworks. PISA is mainly different from TIMSS in that it includes more items relating to the reading, interpretation and evaluation of data, in one form or the other. According to the second set of comparisons given by the external criteria, the main difference is that more items in PISA relates to graphical representations of data. This is a natural consequence when seeking to develop a test that should relate more strongly to realistic or authentic con*texts*. In addition, mathematics in the "algebraic mood" is weighted down in PISA as compared to TIMSS. However, as indicated by the external descriptor "Algebraic expressions", the difference is not mainly related to the use of formal algebraic expressions. The relative stronger emphasis on algebra in TIMSS is therefore related to less stringent notions of algebra as expressed in the concept pre-algebra – items typically setting up tasks were the aim is to find general patterns in sequences of numbers or figures. Finally, Table 3 documents that PISA includes more open-ended response formats than TIMSS.

	Item content descriptor	TIMSS (N=194)	PISA (N=84)
TIMSS Content Domains	Number	29	25
	Algebra	24	8
	Measurement	16	10
	Geometry	16	18
	Data	14	35
	Unclassified		*5*
External criteria	Item-format (SR / SCR / ECR)	66 / 31 / 3	33 / 42 / 25
	Algebraic expressions	16	10
	Calculations	42	37
	Graphics	9	21
	Tables	11	13

Table 3: *Relative distribution of items across content descriptors. The three figures given for item format are Selected Response (SR), Short Constructed Response (SCR) and Extended Constructed Response (ECR) respectively.*

In Table 3 we also find that a few of the PISA items were impossible to classify according to the TIMSS framework. These four items all relate to what could be labelled discrete mathematics, three of which were relating to combinatorics. If these items should be forced into the TIMSS framework, Number would be the most suitable content domain to use.

The two authors of this paper independently placed 73 % of the PISA items into the same content domain. In resolving our disagreement we noted some recurring differences in interpretation. There were two main sources of disagreement. Out of a total of 21 disagreements six related to Geometry vs. Measurement. Many of the problems in PISA relating to Geometry include measurements of lengths, areas or volumes. In the case of reaching consensus for these six items four were classified as Measurement. Furthermore, the two authors disagreed for six items whether they should be classified as Number or Data. These items typically involved the use of tabular or graphical representation of quantities. Five of these ended up being classified as Data.

Table 4 shows in more detail how the PISA content categories, the overarching ideas, match the TIMSS content domains. There is an overall agreement with a rough one-to one correspondence:

- The majority of the Space & Shape items in PISA relate to Geometry in TIMSS.
- The majority of the Quantity items in PISA relate to Number in TIMSS.
- The majority of the Uncertainty items in PISA relate to Data in TIMSS.

			Content Domain						
			Number	Algebra	Measurement	Geometry	Data	Unclassified	Total
Overarching Idea PISA	Space & Shape	Count	1	0	5	14	0	0	20
		% within Overarching Idea	5%	0%	25%	70%	0%	0%	100%
	Change & Relationship	Count	3	6	3	0	10	0	22
		% within Overarching Idea	14%	27%	14%	0%	45%	0%	100%
	Quantity	Count	15	1	0	0	2	4	22
		% within Overarching Idea	68%	5%	0%	0%	9%	18%	100%
	Uncertainty	Count	2	0	0	1	17	0	20
		% within Overarching Idea	10%	0%	0%	5%	85%	0%	100%
Total		Count	21	7	8	15	29	4	84
		% within Overarching Idea	25%	8%	10%	18%	35%	5%	100%

Table 4: *A crosstable showing the correspondence of classification categories when the PISA 2003 mathematics items are classified according to the framework of TIMSS.*

Besides presenting an overall simple correspondence between some of the overarching ideas in PISA and the content domains in TIMSS, Table 4 also show some notable deviations from this pattern. First of all, the overarching idea labelled as Change & Relationship in PISA spreads itself across several of the TIMSS content domains. Phenomena organised under this heading might be described mathematically in several forms of representations. In other words, even if the phenomena might reasonably be put into one class as in PISA, the mathematics involved might vary across the specific representations of these phenomena. The items might relate to simple numerical description of special cases of change or the phenomena might be described as more general relationships, either in algebraic forms or with tables and graphs. Furthermore, it is interesting to note that the overarching idea Space & Shape in PISA also includes several items that match the descriptions of the TIMSS content domain Measurement, a relevant issue that will be returned to in the following discussion.

CONCLUSION AND DISCUSSION

Country specific patterns of achievement

Figures 2 and 3 presented specific profiles of achievements for five countries across the content domains of TIMSS and the overarching ideas in PISA. The main criterion in the selection of these five countries was that each of them represents clusters of countries that previous studies have shown to be stable groupings of countries in both TIMSS 1995 and PISA 2003. In a follow-up of the cluster analysis of the PISA 2003 mathematics achievement profiles Olsen & Grønmo (2006) studied in more detail the characteristic features of each of the profiles. The most prominent findings were that the English-Nordic group had particularly high achievement for items setting realistic and authentic tasks within realistic stimulus materials, and on the other hand these countries perform relatively lower on items requiring exact calculations and use of algebraic expressions, while the group of Central East European countries was documented to have largely the opposite profile, scoring relatively better on purer mathematical tasks requiring of the students to calculate and/or relate to algebraic expressions. Although the profiles for the cluster of Central West European countries and the cluster of East Asian countries also was very distinct, their profile was not equally strongly related to the set of criteria describing the items. The main overall conclusion in that respect is that these figures are consistent with these previous studies. The English and Nordic countries (as represented by Scotland and Norway respectively) have profiles across the content categories that are consistent with the description given above.

Furthermore, it is evident from our detailed comparisons of items in PISA and TIMSS that the main difference between PISA and TIMSS is a relatively stronger emphasis in PISA on realistic contexts. In other words, the differences between PISA and TIMSS items are very much related to the relative strengths and weaknesses for the English-Nordic and East European profiles with the effect that the English-Nordic countries seems to be relatively more successful in mathematics as it is defined in PISA than in TIMSS, and vice versa. East Asian countries and also other clusters of countries did not have such articulated differences along the content domains, and consistent with this, many of these countries did not shift so much in the ranks in the two studies.

Operational sharpness of content domains

Our initial analysis of the TIMSS and PISA framework suggested that TIMSS has a much more fine-grained operational definition of the content dimension than PISA. This is supported by the data in Figures 2 and 3 where it is seen that the overall variation for specific countries across the five content domains in TIMSS is larger than the corresponding variation across the four overarching ideas in PISA. It is reasonable also to think that the relatively larger between-country variation seen in TIMSS is related to this increased operational sharpness. This sharpness of definition has two consequences. First of all, data from comparative studies like PISA and TIMSS should highlight differences between countries. Such differences, either between diverging educational systems or between educational systems that are closely related, are useful starting points for educational research (Olsen, 2005). Thus, from this perspective, enlarging the differences between countries would increase the potential for meaningful comparison. On the other hand, seen from a measurement perspective, increased differential performance across the items or across the content domains is also a potential threat for meaningful comparison since this can indicate potential international measurement errors due to content-by-country interactions (Wolfe, 1999). Put simply, the effect of replacing specific items or by redefining the relative weights of each content domain would likely have a larger effect on the total score in mathematics in TIMSS than in PISA. Thus, we could say that TIMSS gives us a measure of mathematics achievement that is sharper than PISA, with the possible side effect that the measure is less stable.

The two authors started out to independently categorise the PISA items from 2003 using the framework of TIMSS 2003 with an agreement of 73%. We believe that the main cause of the problem of non-perfect classification agreement is the fact that these items truly are crossovers of these domains. It is likely better to interpret this index of classification agreement as a measure of the degree to which it is possible to link PISA items to the TIMSS framework, than as a measure of marker reliability. Furthermore, the disagreement seemed to be rather systematic, pointing to what we have interpreted at indications that the TIMSS 2003 framework (Mullis et al., 2001) in some cases is a little unclear. The topic Measurement and its relation to the topics Geometry seemed problematic; the same with Numbers versus Data. Discussing these disagreements we saw that according to the framework, it was good arguments for both types of classifications. This is equally true for a number of items were we happened to agree. In this respect it is interesting to note that the new framework for TIMSS 2007 has deleted the category Measurements. Some of the items originally classified as Measurement in 2003 will also be used as link items in the 2007 study. Accordingly, these items had to be reclassified, and several of the items previously categorised as Measurement are categorised as Geometry items in 2007. Reasons given for this change in the TIMSS framework so far has been rather technical, pointing out that in trend studies we can not have that many categories to be able to report measures of trends with the appropriate quality. Our analysis adds substantial arguments for why it is wise to exclude the Measurement category: By doing so the quality of the reporting scales for the remaining four content domains will increase due to the increased specificity in the operational definitions of the remaining content domains.

The case of pure and applied mathematics in school

For many countries it is a goal that all students after compulsory school should have a type of competence we may call mathematical literacy, that they are well prepared to solve daily life problems using mathematics and can be active citizens in a modern society. PISA seems well suited to answer if students in a country have reached this goal. TIMSS complements this information. Referring back to the model of the mathematization cycle in Figure 1, PISA measures the mastery of all the processes involved in solving mathematical problems originating from a real world context. TIMSS on the other hand gives a measure of the mastery of the mathematical processes in the right-hand side of figure 1. Taken together these studies can therefore identify more specifically how the mathematical literacy may be increased in a country: If a country or group of countries achieve better in PISA than in TIMSS, it may indicate that students have problems with their competence in pure mathematics, in general or in specific topics in mathematics. If the opposite is the case, a country is achieving better in TIMSS than in PISA, it may indicate that students are not often presented for the full mathematical cycle for applied mathematics as illustrated in Figure 1.

We have documented that a country like Russia, representing the East European profile, do relatively better in TIMSS than in PISA. From the simple proposition suggested above the interpretation would be that most of the East European countries give little attention to the left-hand side of the mathematization cycle. The general message that this example serves to communicate is that “back to basics” is not a complete solution if the aim is to foster students with mathematical literacy.

Japan, representing the East-Asian profile, is high achieving in both studies, more pronounced in TIMSS than in PISA. This may indicate that mathematics in school in the East Asian countries to a great extent focus on pure mathematics in all topics, while at the same time they also give some attention to the full cycle of applied mathematics. The Netherlands is, as Japan, among the high achieving countries in both studies. Nevertheless, there are some distinct differences between their achievement levels in different topics in mathematics revealed in the TIMSS study. While Japan and The Netherlands achieve equally well in the topics Number, Measurement and Data in TIMSS, there is a clear difference between their achievement levels in Algebra and Geometry. This tells us that even high achieving countries may not be identical when it comes to what they focus in curriculum. Algebra and Geometry seems to be much more in focus in Japan than in The Netherlands. And when it comes to achieving well in mathematical literacy, as tested in PISA, The Netherlands is doing equally well compared to Japan. We take this to indicate that the "basics" of most importance are the fundamental concepts of numbers and operations with numbers, more so than a "basics" expressed as formal insight into geometry and algebra.

The shape of the graphs for achievement levels in Norway and The Netherlands in specific topics in mathematics in TIMSS, display that they are more or less the same. This indicates that the mathematics curriculum in school in both countries have many similarities. The difference, however, is that in opposition to Norway, The Netherlands is a high achieving country in general, in TIMSS and even more pronounced in PISA. Norway is achieving lower than any of the other Nordic countries in PISA, and even lower in TIMSS at both grade levels. When we compare the Norwegian students' achievement in TIMSS with what is focused in the Norwegian curriculum in mathematics, the topic in TIMSS which seems most problematic is Number. That the achievement in Algebra is even lower is easy to explain, since this topic, to a great extent, has become a very small part of what is taught in compulsory school for the last decade. Number however, is an extensive part of the curriculum all trough compulsory school. That Norwegian students lack elementary knowledge and skills in Number, was even more pronounced for students in grade 4 in TIMSS, but we have not presented this result in this article. Since Number is an extensive part of the curriculum in compulsory schools in Norway, as in most other countries, we will point to what we have referred to earlier in this article. As a consequence of a growing focus on applied mathematics, it may be a problem if too little attention is given to the pure mathematics, especially pure mathematics as elementary knowledge and skills in a topic as Number. The Norwegian results in PISA and TIMSS seem to exemplify that this may be the case in Norwegian schools. Earlier analysis of what kind of items in PISA Norwegian students performed well and not well on also underline that it is a reasonable interpretation of the results. While Norwegian students performed relatively good on PISA items close to what students may expect to meet in their daily life, they performed low on items acquiring any type of exact calculations (Olsen & Grønmo, 2006). Basic skills in elementary mathematics seem to be necessary conditions for doing well in applied mathematics as tested in PISA. This is also supported by the result in TIMSS, countries doing well on items in problem solving is also achieving well on more elementary items (Mullis et al. 2004).

Our analysis and comparisons between TIMSS and PISA support that in order to do well in daily life mathematics; students need a basis of knowledge and skills in pure mathematics, especially elementary knowledge and skills in numbers. This indicates that it is important in school curriculum that mathematical literacy is not seen as an alternative to pure mathematics. A reasonably high level of competence in pure mathematics seems to be necessary for any type of applied mathematics, as we pointed out in our discussion of Figure 1.On the other hand, if too little attention is given to the full cycle of applied mathematics, it is unlikely that students will develop the type of competence we may call mathematics literacy.

REFERENCES:

Bjørkquist, O. (2001). Matematisk problemløsing (H. Strømsnes, Trans.). In B. Grevholm (Ed.), *Matematikk for skolen* (pp. 51-70). Bergen: Fagbokforlaget.

Ernest, P. (1991). *The philosophy of mathematics education*. London: Falmer Press.

Gardiner, A. (2004, June). *What is mathematical literacy?* Paper presented at the ICME 10, København.

Grønmo, L. S., Bergem, O. K., Kjærnsli, M., Lie, S., & Turmo, A. (2004). *Hva i all verden har skjedd i realfagene? Norske elevers prestasjoner i matematikk og naturfag i TIMSS 2003*. Oslo: Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.

Grønmo, L. S., Kjærnsli, M., & Lie, S. (2004). Looking for cultural and geographical factors in patterns of response to TIMSS items. In C. Papanastasiou (Ed.), *Proceedings of the IRC-2004 TIMSS* (Vol. 1, pp. 99-112). Nicosia: Cyprus University Press.

Kjærnsli, M., Lie, S., Olsen, R. V., Roe, A., & Turmo, A. (2004). *Rett spor eller ville veier? Norske elevers prestasjoner i matematikk, naturfag og lesing i PISA 2003*. Oslo: Universitetsforlaget.

Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 International Science Report. Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Boston: TIMSS & PIRLS International Study Center, Lynchs School of Education, Boston College.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2005). *IEA's TIMSS 2003 International Report on Achievement in the Mathematics Cognitive Domains: Findings from a Developmental Project*. Boston: TIMSS & PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 International Mathematics Report. Findings From IEAs Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Boston: TIMSS & PIRLS International Study Center, Lynchs School of Education, Boston College.

Mullis, I. V. S., Martin, M. O., Smith, T. A., Garden, R. A., Gregory, K. D., Gonzalez, E. J., et al. (2001). *TIMSS Assessment Frameworks and Specifications 2003*. Boston: International Study Center, Lynch School of Education, Boston College.

National Center for Education Statistics (NCES). (2005). Comparing NAEP, TIMSS, and PISA in Mathematics and Science. Retrieved May 24th, 2005, from http://nces.ed.gov/timss/pdf/naep_timss_pisa_comp.pdf

NCTM (National Council of Teachers of Mathematics). (1989). *Curriculum and evaluation standards for school mathematics*. Reston: NCTM.

Neidorf, T. S., Binkley, M., Gattis, K., & Nohara, D. (forthcoming). *A Content Comparison of the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 Mathematics Assessments (NCES 2006-029)*. Washington, DC: National Center for Education Statistics.

Nohara, D. (2001). A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA). Working Paper No. 2001-07. Retrieved May 24th, 2005, from http://nces.ed.gov/pubs2001/200107.pdf

OECD. (2003). *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*. Paris: OECD Publications.

OECD. (2004). *Learning for Tomorrow's World. First Results From PISA 2003*. Paris: OECD Publications.

Olsen, R. V. (2005). *Achievement tests from an item perspective. An exploration of single item data from the PISA and TIMSS studies, and how such data can inform us about students' knowledge and thinking in science.* Oslo: Unipub forlag.

Olsen, R. V. (2006). A Nordic Profile of Mathematics Achievement: Myth or Reality? In J. Mejding & A. Roe (Eds.), *Northern Lights on PISA 2003 - a reflection from the Nordic countries* (pp. 33-45). Copenhagen: Nordic Council of Ministers.

Olsen, R. V., & Grønmo, L. S. (2006). What are the Characteristics of the Nordic Profile in Mathematical Literacy? In J. Mejding & A. Roe (Eds.), *Northern Lights on PISA 2003 - a reflection from the Nordic countries* (pp. 47-57). Copenhagen: Nordic Council of Ministers.

Schoenfeld, A. H. (1992). Learning to Think Mathematically: Problem Solving, Metacognition, and Sense Making in Mathematics. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 334-370). New York: MacMillan.

Schoenfeld, A. H. (2004). The Math Wars. *Educational Policy, 18*(1), 253-287.

Skovsmose, O. (1994). *Towards a philosophy of critical mathematics education*. Dordrecht: Kluwer Academic Publishers.

Wolfe, R. G. (1999). Measurement Obstacles to International Comparisons and the Need for Regional Design and Analysis in Mathematics Surveys. In G. Kaiser, E. Luna & I. Huntley (Eds.), *International Comparisons in Mathematics Education*. London: Falmer Press.

Zabulionis, A. (2001). Similarity of Mathematics and Science Achievement of Various Nations. *Education Policy Analysis Archives, 9*(33).