

Comparison of Mental Structures of Eighth-graders in Different Countries on the Basis of Fennema-Sherman Test

Jari Metsämuuronen¹

¹ Faculty of Behavioral Sciences, Helsinki University, Finland

Correspondence: Jari Metsämuuronen, Faculty of Behavioral Sciences, Helsinki University, Siltavuorenpenger 5 A, P.O.B. 9, FIN-00014, Finland. E-mail: Jari.Metsamuuronen@methelp.com

Received: May 21, 2012 Accepted: May 30, 2012 Online Published: November 12, 2012

doi:10.5539/ijps.v4n4p1

URL: <http://dx.doi.org/10.5539/ijps.v4n4p1>

Abstract

The intended factor structure of attitude scales in TIMSS 2007 dataset ($N > 240,000$) is totally unstructured in several countries. There are two inseparable phenomena which explain the differences between countries. First, the ability level of the students is strictly connected with the level of maturation in thinking. Second, there are differences between countries which cannot be explained by the differences in achievement level. Four distinguishable structures are identified: “Western” structure, “African–Asian” structure, “Middle Eastern–Asian” structure, and “high-performing East Asian” structure.

Keywords: student attitudes, mathematics achievement, international comparison, foreign countries, achievement tests, attitude measures, cross cultural studies, cultural differences

1. Introduction

International comparisons have evidently shown that mathematical achievement is positively associated with attitudes toward mathematics and mathematical self-concept (e.g., Yoshino, 2012; Areepattamannil, 2012; House & Telese, 2008; Shen & Tam, 2008; Eklöf, 2007; House, 2006a; 2006b; Van den Broeck, Opdenakker & Van Damme, 2005; Ma & Xu, 2004; Hammouri, 2004; Webster & Fisher, 2000; Papanastasiou, 2000a; Lokan & Greenwood, 2000; Gadalla, 1999; Ma & Kishor, 1997a; 1997b; Hembree, 1992). However, several authors have noted—though not necessarily deepened the discussion—that there are notable *differences* between the countries when it comes to the intensity of the association between the concepts (e.g., Yoshino, 2012; House & Telese, 2008; Kadjevich, 2006; 2008; Wilkins, 2004; Shen, 2002; Papanastasiou, 2002; 2000b; Leung, 2002; Stevenson, 1998). For example, Wilkins (2004) found a positive relationship between eighth-graders’ achievement and self-concept for *almost* all countries that participated in Trends in International Mathematics and Science Study (TIMSS) 1995, and Shen (2002) reported a positive relationship in *almost* all countries that participated in TIMSS 1999. Kadjevich (2008) noted that in *almost* all countries, each dimension of mathematics attitudes was positively related to mathematics achievement in TIMSS 2003. Remarkable differences in correlations are reported between some countries: from 0 (0.02) to 0.61 in Indonesia and Korea (Shen, 2002), from 0 (0.16) to 0.46 in Moldova and Chinese Taipei (Shen 2002), from 0 (0.00) to 0.40 in Macedonia and Korea (Kadjevich, 2008), and from 0 (–0.02) to 0.46 in the Philippines and Korea (Wilkins, 2004). Kadjevich (2008) shows that in countries with higher mathematics achievement, the relationship between mathematics self-concept and mathematics achievement is stronger. He thus suggests that countries with more demanding mathematics curriculum have a stronger relationship between mathematics achievement and mathematics attitude dimensions. Additionally, Else-Quest, Hyde and Linn (2010) reported that there are quite remarkable differences between countries when it comes to difference between boys’ and girls’ attitudes towards mathematics.

The differences between the countries are such remarkable that some writers have deepened the analysis on the TIMSS 2007 dataset from this respect. Metsämuuronen (2012) shows that the shortened version of Fennema–Sherman Mathematics Attitude Scales (Fennema & Sherman, 1976) used in TIMSS 2007 (Mullis, Martin, & Foy, 2008) and its predecessors 1995, 1999, and 2003 as well as in Programme for International Student Assessment (PISA), carries severe weakness when it comes to the international comparisons. Namely, two items, “*Mathematics is not one of my strengths*” and “*Mathematics is more difficult for me than for many of my classmates*”, are shown to be too difficult for the students at the lowest achievement level. They also seem to

carry culturally sensitive elements though Metsämuuronen do not elaborate them. Bouhlila (2011) ponders the reasons behind the differences in achievement from the cultural viewpoint and finds the students' attitude as one of the factors behind the poor results in the Middle East and North African (MENA) countries. Liou (2010) notes that 49 participating countries can be formed into six groups; countries within a cluster tend to have geographical proximity or shared cultural and educational backgrounds.

This article deepens the treatment of Metsämuuronen (2012) and focuses on the cultural sensitive elements in the Fennema-Sherman test: the reanalysis of the same dataset strongly suggests that there are remarkable differences in *thinking* between the students from different countries. The article utilizes the finding of Metsämuuronen that in several countries the expected factor structure of Fennema-Sherman test is fragmented either into "totally unstructured" or "moderately unstructured" factor structure and 1) elaborates the finding that there is a strict connection between the achievement level of the students and the fragmentation of the factor structure, 2) shows that the cultural matters are connected with the unexpected factor structure, and 3) that different cultural and geographical areas seem to share common "unexpected" factor structures.

2. Data and Methods

2.1 Data

As is usual in the high level international testing settings – such as in TIMSS and PISA – the development of the test instruments, their translations and country-wise adaptations are well reported (see the basic reports of Olsen, Martin & Mullis, 2008, for TIMSS 2007 and OECD, 2009, for PISA 2006 and OECD, 2012, for PISA 2009). Especially for TIMSS 2007, Johansone and Malak (2008) have described the translations and country-wise adaptations. The system of producing internationally comparable results is very heavy and hence, the tests, background questionnaires, data processing and the data itself in TIMSS processes can be taken as the top quality "products" in the area of student achievement. The dataset is available in Foy & Olson (2009).

Two sets of information are relevant from the massive datasets: Mathematics achievement estimates and Fennema-Sherman test for attitudes in Mathematics. In order to produce comparable results with Metsämuuronen (2012), the same data and procedures are used. All the countries in TIMSS 2007 ($N = 57$) are combined into one dataset consisting of a total of 248,160 8th grader students. For analysis, the dataset is divided into 20 percentiles ($N \approx 12,000$ in each) on the basis of the first plausible value of Mathematics achievement (Table 3). After dividing the data into percentiles it is worthwhile to note that the range in the lowest and highest percentile is much wider than with other groups because of representing the tail populations and that none on the percentile shows Normal distribution; in percentiles 2–19 the population is merely uniform than normally distributed. There is thus some estimation error in the parameters of Explorative Factor Analysis (EFA) used in the analysis procedure. However, because of the large sample sizes the results can be taken stable. Naturally, the results are restricted to 8th graders.

2.2 Fennema-Sherman Test

The original (or latest) version of the Fennema-Sherman Mathematics Attitude Scales (Fennema & Sherman, 1976) includes nine dimensions (see Vezeau *et al.*, 1998; Alkhateeb, 2004; compare Melancon, Thompson, & Becnel, 1994 who identified eight dimensions, Mulhern & Rae, 1998 who identified six dimensions, and Tapia & Marsh, 2005, who used four dimensions and Liou, 2010 who sees dimensions of intrinsic- and extrinsic motivations and self-concept there). Usually in the internationally setting the shortened Fennema-Sherman Mathematics attitude scale is divided into two sets of questions (in TIMSS 2007 questions 8 and 9) with the same question: "*How much do you agree with these statements about learning mathematics?*" The statements in Question 8 are as follows:

- a. I usually do well in mathematics,
- b. I would like to take more mathematics in school,
- c. * Mathematics is more difficult for me than for many of my classmates,
- d. I enjoy learning mathematics,
- e. * Mathematics is not one of my strengths,
- f. I learn things quickly in mathematics,
- g. * Mathematics is boring,
- h. * I hate mathematics.

The items with the asterisk (*) are opposite to the scale, and thus they are reversed before scoring. The last question in the set "I hate mathematics" was originally negative but the item was reversed ("I like mathematics") before releasing the data. For the analysis it was reversed back to reach the original structure of the test: two

negative and two positive items for both Dimension 1 and 2. The statements in Question 9 are as follows:

- a. I think learning mathematics will help me in my daily life,
- b. I need mathematics to learn other school subjects,
- c. I need to do well in mathematics to get into the <university> of my choice,
- d. I need to do well in mathematics to get the job I want.

In what follows, shortened versions of the items are used to shorten the narrative, and texts on tables and figures. The abridged versions are obviously recognized. As seen on Table 1, three dimensions are constructed as follows:

- 1) Liking MATH: Question 8, items b, d, g*, and h*,
- 2) Self-concept in MATH: Question 8, items a, c*, e*, and f, and
- 3) Experiencing utility in MATH: Question 9, items a to d.

Alpha reliabilities for the scales are respectively .72, .70, and .74 in the whole dataset. Metsämuuronen (2012) showed that alpha reliabilities for the scale “Self-Concept in MATH” are very low (below 0.60) in the lowest achievement groups because of two complex negative items on the test.

2.3 Statistical Methods

The shortened Fennema-Sherman test is explored two ways with EFA. First, the analysis is done country-wise with the Principal Axis Factoring (PAF), three factor solution, and Promax rotation with Kaiser Normalization. The factor model in each country is categorized on the basis of the expected factor structure and the variation explained by this three factor solution is extracted. Second, the achievement level of the students is divided into 20 percentiles ($n \approx 12,000$ in each percentile), the factor structure is categorized, and factor loadings for indicator items are collected at each achievement level. The qualitative categorization of the factor structures and the quantitative indicators are used to profiling the different countries and cultural areas into homogeneous sets.

During the analysis, the alpha reliability (Kuder& Richardson, 1937; Gulliksen, 1950; Cronbach, 1951) is used as an indicator for the overall reliability in the attitude score. Though it is known that alpha reliability always underestimates real reliability (Gulliksen, 1950; Lord & Novick, 1968; Tarkkonen, 1987; Vehkalahti, 2000) in practice it is the most used indicator for reliability (Hogan, Benjamin, & Brezinski, 2000).

The Classic Pearson product-moment correlation coefficient is used when assessing the association between the negative- and positive counterparts of the items.

3. Results

3.1 Expected and Unexpected Factor Structure of the Fennema-Sherman test

The structure in the shortened Fennema-Sherman test is simple: four items on each dimension (see Table 1.) and two negative items in each of the first two dimensions as seen also in Metsämuuronen (2012). This kind of *Expected factor structure* can be found in all European countries (except in Bulgaria and Romania), Australia, Canada, Israel, the United States, and Russia.

Table 1. (Expected) factor structure in the USA in TIMSS 2007

	PatternMatrix ^a		
	1	2	3
Math is more difficult to me than many of my classmates	-0.775 ^b		
Math is not one of my strengths	-0.749		
Usually I do well in Math	0.742		
I learn things quickly in Math	0.710		
I like Math		0.855	
I enjoy learning mathematics		0.853	
Math is boring		-0.743	

I would like to take more Math	0.616
I need Math to get the job I want	0.70
I need math to get into the <UNIVERSITY>	0.67
I need Math to learn other subjects	0.58
Mathematics will help in my daily life	0.55
Extraction Method: Principal Axis Factoring.	
a. Rotation converged in 5 iterations.	
b. Values >.30 are shown	

After performing exploratory factor analysis (EFA) separately in all countries, it is notable that in several countries, this structure cannot be found. Several countries share an unexpected “*Totally unstructured*” factor structure, as seen in Table 2—here Indonesia as an example: the negative items correlate with each other more than with their expected positive counterparts, and thus they produce their own factor of “negative items.” The countries with “*Totally unstructured*” factor structure are, in the alphabetic order, as follows: Armenia, Dzac, Egypt, Ghana, Indonesia, Jordanian, Kuwait, Lebanon, Morocco, Mongolia, Malaysia, Oman, Palestine, Qatar, Saudi-Arabia, Syria, Thailand, and Tunisia. The third structure, “*Moderately unstructured*” factor structure (see Metsämuuronen 2012), can be found in Bahrain, Bulgaria, Georgia, Iran, Japan, Korea, Romania, Singapore, and Turkey. This structure is characterized by some minor impurities in the expected factor structure. “*Unique structure*”, seen in Botswana and El Salvador, is characterized by a unique logic in the factor structure.

Table 2. (Unexpected) Factor structure in Indonesia in TIMSS 2007

	PatternMatrix ^a		
	Factor		
	1	2	3
I like Math	0.790 ^b		
I learn things quickly in Math	0.673		
I enjoy learning mathematics	0.641		
Usually I do well in Math	0.619		
I would like to take more Math	0.472		
I need Math to get the job I want		0.713	
I need math to get into the <UNIVERSITY>		0.692	
Mathematics will help in my daily life		0.435	
I need Math to learn other subjects		0.363	
Math is more difficult to me than many of my			0.658
Math is not one of my strengths			0.464
Math is boring			0.434
Extraction Method: Principal Axis Factoring.			
a. Rotation converged in 5 iterations.			
b. Values >.30 are shown			

Metsämuuronen (2012) noted that many countries sharing a common “unexpected” structure are close neighbors and share common values and culture. Coincident maybe, but seems to be so that with two exceptions (Mongolia and Thailand), all countries with “*Totally unstructured*” factor structure share an Islamic heritage. Several countries in East Asia seem to share the “*Moderately unstructured*” factor structure. In what follows, this common structure of thinking is called “mental structure” – Magoroh Maruyama’s (1980) concept “mindscape” could have been used (see also Caley & Sawada, 1994). Fritz (2006) compresses the idea of mindscape as follows: “*a mindscape is a way of thinking that affects all our thoughts. It is a strategy that guides the way we*

select our objectives and how we act to reach them. A mindspace is one or a collection of a few of the most general response rules we have.” Heuristically, the concept of “mindspace” seems to explain the differences between the different cultures and similarities of certain areas – the “East Asian structure” seems to differ from the “European structure”.

3.2 Connection of Achievement and Fragmentation of the Structure of the Attitude Test

It is quite an interesting fact that the country-wise discrepancy level of the expected structure, measured as the percentage of explained student variance in EFA, and the achievement level in TIMSS 2007 mathematics score, estimated by the unweighted first plausible value of MATH achievement in TIMSS 2007 dataset, correlate remarkably ($r = 0.78$, see Fig. 1); the more the discrepancy, the less the achievement. This seems to confirm the results of Kadijevich (2006, 2008): the higher the mathematics achievement, the stronger the relationship between mathematics self-concept and mathematics achievement. However, it is good to note that the logic is not flawless: the model of three factors might quite well explain the mental structure of the students, even though the structure would not be as expected. This is evident with the countries with “moderately unstructured” factor structure and simultaneously high performance in the TIMSS 2007 test (Korea, Singapore, and Japan) indicated with the darker far right-handed circles in Figure 1.

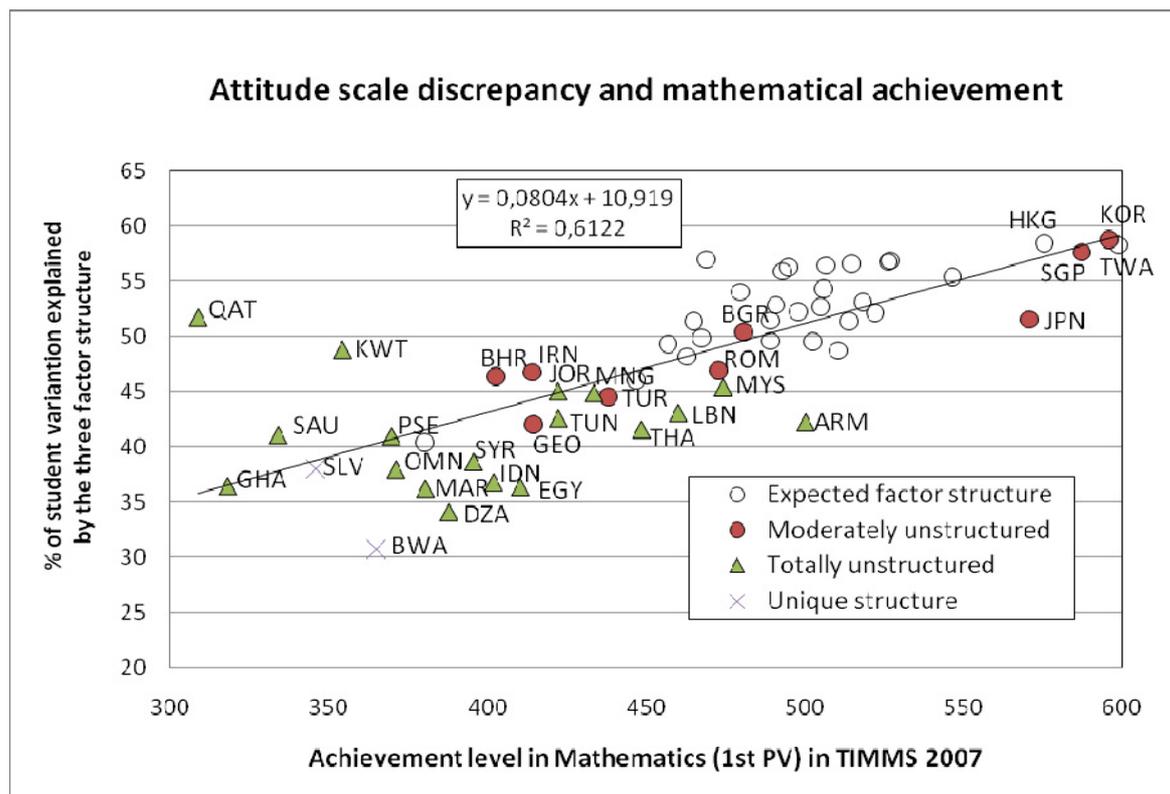


Figure 1. Correlation between attitude structure discrepancy and achievement level in TIMSS 2007 mathematic test

Figure 1 and the high correlation of discrepancy in factor structure and achievement level raises question of whether is it possible to explain the discrepancy in attitude structure by the achievement level, or are the differences between the achievement levels caused by the different mental processes of students? Both possibilities are tackled in the following Section. Figure 1 is elaborated and re-used later with Figure 5.

3.3 Mental Structure as Differences in the Achievement Levels?

Students' factor structures depend strictly on the students' achievement levels (see Table 3, and Figure. 2)—as suggested also by Kadijevich (2006; 2008) and Metsämuuronen (2012). The maturation of the thinking can be divided into four phases clearly seen in Table 3. The phases are named and argued as (1) Concrete, (2) Developing, (3) Formed, and (4) Matured in Metsämuuronen (2012).

At the lowest level of achievement, that is, in the percentiles 1–3 where the achievement level is less than 338 points in the TIMSS mathematics scale, the factor structure is characterized by *one clear factor of negative items*. Technically speaking, at this level all the negative items—except “I HATE MATHEMATICS”—correlate each other more than with the expected positive counterparts. The loadings of the negative items are highly positive. At Table 3, it can be seen that the absolute value of the bivariate correlation coefficient of variable “MATHS IS NOT ONE OF MY STRENGTHS” and its positive counterpart “I USUALLY DO WELL IN MATHS” is $r = 0.12$ or less, and for variables “MATH IS BORING” and “I ENJOY LEARNING MATH” it is $r = 0.30$ or less. This kind of technical reason also affects the reliability of the scores to be low. In the case, the reliabilities in the lowest percentiles for the score of “self-concept in math” are less than $\alpha = 0.50$ (Table 3), which classically is regarded as too low reliability for a test (see Knapp & Brown 1995). It may be worth pointing out that during the analysis it was noted that the boundaries of the levels depend on what countries were selected for the analysis. Therefore it is recommended to take the boundaries *substantive*, say, with these countries and with this sample the boundaries are as described here.

Table 3. Achievement levels of the students and single indicators of mental structures

	Percentile (N > 12000 in each group)	Achievement mean in MATH ¹⁾	Correlation between "BORING" and "ENJOY" ²⁾	Correlation between "WEAK" and "I DO WELL" ³⁾	Reliability of the scale "Self-Concept in math" ⁴⁾
4) Matured	20	675	-.636	-.666	0.858
	19	611	-.633	-.652	0.854
	18	581	-.629	-.625	0.842
	17	558	-.622	-.573	0.818
	16	540	-.633	-.544	0.805
	15	524	-.616	-.533	0.799
	14	509	-.611	-.505	0.781
	13	494	-.605	-.477	0.773
	12	480	-.592	-.461	0.753
	11	466	-.589	-.455	0.753
3) Formed	10	452	-.551	-.421	0.731
	9	437	-.535	-.373	0.707
	8	422	-.519	-.352	0.693
2) Developing	7	406	-.502	-.316	0.666
	6	389	-.456	-.271	0.641
	5	371	-.428	-.235	0.609
1) Concrete	4	350	-.374	-.177	0.552
	3	325	-.305	-.120	0.498
	2	292	-.258	-.053	0.417
	1	222	-.149	.058	0.259

1) Unweighted, 1st Plausible value (PV) in TIMSS 2007 dataset

2) variables “MATH IS BORING” and “I ENJOY LEARNING MATH”

3) variables “MATHS IS NOT ONE OF MY STRENGTHS” and “I USUALLY DO WELL IN MATHS”

4) Variables “MATH IS MORE DIFFICULT FOR ME THAN FOR MANY OF MY CLASSMATES” (reversed), “I LEARN THINGS QUICKLY IN MATH”, “MATHS IS NOT ONE OF MY STRENGTHS” (reversed), and “I USUALLY DO WELL IN MATHS”.

It may be worth noting a tiny detail in Table 3: In the dataset it is possible to predict the rough achievement level of the students very accurately, only by knowing the correlation between two variables. Namely, the correlation of variables “MATHS IS NOT ONE OF MY STRENGTHS” and “I USUALLY DO WELL IN MATHS” correlates practically one-to-one with the achievement group mean: the correlation between the correlation and percentile

group mean is $r = -0.98$. Hence, from the technical point of view there is no need to take any achievement tests—there is a need only to know this bivariate correlation for a student to get a rough estimation of the students' ability level! Of course, more than a serious idea, this detail conveys the direct association of achievement level and mental capability. There are, most probably, differences between countries when it comes to the strength of the correlation. The set of questions and the set of countries in the dataset may also influence the results. Here, the subject is not discussed in-depth.

The critical question is: how it is possible that there is practically no correlation between two items obviously measuring the same thing, but with opposing wordings? The same question can be put in another form: why do many students at the lowest ability level answer illogically to the negative item? Of course, there may be several good reasons behind the phenomenon discussed, for example, in Metsämuuronen (2012). One of the plausible reasons may be that the *low-ability level students' general ability to read is so low that they just do not understand the question*. This may also explain their poor mathematics skills: they were not able to read the stems in the questions. The test takers, however, are eighth-graders; say, they have studied in the formal educational system for eight years. If, after this much of education, they cannot read some simple questions, our educational systems all over the world at the moment produce several tens of thousands of failures. Anyhow, this should be taken as a serious explanation for the problem.

Another reason, elaborated more here, may be that at the lowest level of achievement, the students' *level of abstract thinking may be low*: many of the lowest ability level test takers may react to a negative wording in too concrete a way. When the students with this low ability level see the negative sentence "MATH IS BORING", quite many of them may think: "*Oh, no, math is not boring—I like mathematics!*", and select the *positive* alternative instead of the required *negative* alternative. This seems to be the logic with small children when answering a Likert-type of scale with sad and happy faces: they may have enough general ability to understand the positive sentence and react adequately, but not enough abstract level thinking, first, to understand the relevancy of the negative wording and, second, to judge whether they have a positive or a negative opinion of this negative sentence. And actually it is quite a complicated mental process to arrive at such a judgment; it is not always easy to judge what would be the *positive* meaning of a *negative* item. For this reason, the lowest level mental phase is called "*Concrete level*". Metsämuuronen (2012) also discusses the different levels of complexity in attitude test items – they are not addressed here.

The highest level of abstract thinking is characterized by the expected factor structure: the higher the achievement levels, the higher loadings and correlations between the corresponding variables in the expected factor. Hence, the label is "*Matured level*" of thinking. This level includes percentiles 10–20 and it requires around score 445 or more in the TIMSS mathematic scale to achieve this level of abstract thinking. Reliabilities for the score of "self-concept in math" range $\alpha = 0.73$ – 0.86 .

The in-between categories are anchored to these extremes. The second lowest level, *Developing level*, is technically characterized by the fact that variable "MATH IS BORING" has a growing loading in the *correct factor*—and thus, the intended factor structure has started to develop. However, the other negative variables are still correlating positively with each other without corresponding positive variables in the same factor. This level of abstract thinking is present in the percentiles 4–6 with scores ranging 338–397 in the TIMSS mathematics scale. Reliabilities for the score of "self-concept in math" range $\alpha = 0.55$ – 0.64 which shows very low accuracy for the test. The second highest level is called *Formed level*, because all the factors are formed but they are still immature. Technically speaking, the factor "Liking math" is formed as it should be. However, the other factor with negative items, "Self-concept in math" is characterized by negative loadings for the positive items and positive loadings for the negative items; hence, the negative items are dominating the factor loadings. This level of thinking is present in the percentiles 7–9 with scores ranging 398–444 in the TIMSS mathematic scale. Reliabilities for the score of "Self-concept in math" in these percentiles range $\alpha = 0.67$ – 0.71 .

These four levels of thinking—Concrete, Developing, Formed, and Matured—seem to be obvious and clear with a large number of international students. The levels form strict continuity, as presented in Fig. 2. When giving the phases a numeric rank order from 1 (Concrete) to 4 (Matured), the mean rank in the countries is 2.9, indicating that on average the students draw near the Level 3, that is, they are at the "Formed" level of thinking. On the basis of 20 percentiles, 55% of the students are at the "Matured level", 15% at the "Formed" level, 15% at the "Developing" level, and as many as 15% of the students are at the "Concrete" level of thinking.

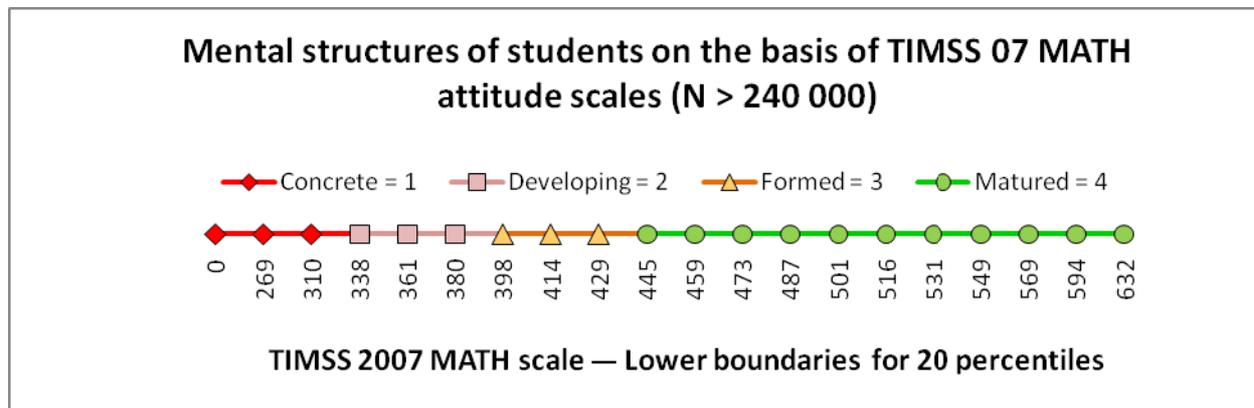


Figure 2. Mental structures of students on the basis of TIMSS 2007 MATH attitude scales

3.4 Comparison of Mental Structures of Students in Selected Countries —Does the Cultural Background Explain the Phenomenon?

On the basis of the analysis above, it seems evident that the mental structure of students associates with their achievement level. The numeric analysis of the combined mass of international students shows that the higher the achievement level, the more matured the thinking. This is not, however, the whole story. Namely, there are drastic differences between the countries concerning the mental structures of eighth-graders' thinking (see Figures 3 and 4).

In Figures 3 and 4, differences between non-European Western and East Asian profiles of mental structure can clearly be seen. When comparing two “Western” countries, e.g., the United States and Russia, there is actually no difference in the profiles, except that in the United States there is more variability in the lowest part of the scale (Figure 3). One easily notices that, first, in both cases the overall structure follows the international structure: the more the achievement, the more the matured thinking. Second, in both cases the “Matured” phase of thinking requires as much achievement as the international structure suggests: in the international data the minimum achievement level to reach the highest mental level is around 445 points in the TIMSS scale, whereas in the United States the students reach the matured level at around 430 points, and in Russia at around 445 points. Though the profiles of students in Canada and in Australia are not as textbook-like as in the United States and Russia the mean rank of the profile shows that on average the students are much nearer the “Matured” level than is the international mean rank. The profiles of Ukraine and Georgia show that quite many students at the middle range achievement level do not reach the most matured phase of thinking—in many higher percentile groups, the factor structures show “Formed” structure rather than “Matured” structure. The comparison leads to a question as to whether it is possible that in some countries with the similar cultural backgrounds the educational system produces *better* mental qualities (note that in what follows the word “*different*” is used). The question will be discussed in what follows in Section 4.

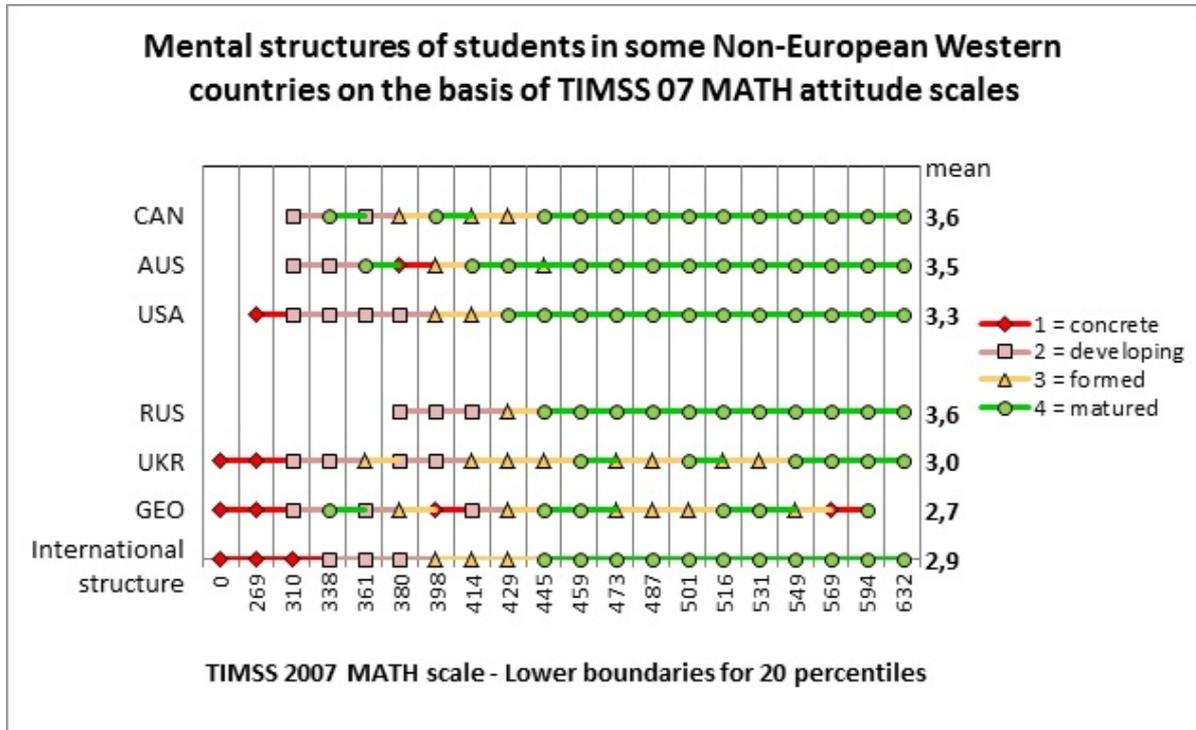


Figure 3. Mental structures of students in some Non-European Western countries on the basis of TIMSS 2007 MATH attitude scales

In Figure 4, one sees a dramatic difference compared with the Western profile seen in Figure 3: in the countries from East Asia—except in Korea and Japan—the achievement level has nothing to do with the “maturation” of thinking measured with the attitude scale discrepancy. These interesting Asian countries in TIMSS 2007 are Indonesia, Thailand, Mongolia, Malaysia, Taiwan, or Hong Kong, where even the best students show “concrete” structure in thinking. Especially in Indonesia and Thailand, there is no such achievement level percentile group where the fully expected factor structure would be present. The fact raises several questions of real cultural differences—maybe even “cultural mindscapes” (see Maruyama 1980)—behind the phenomenon or at least *exacerbated-* or *inseparable linkage* of achievement level and cultural effects in explaining the phenomenon. These are discussed in the final section. Another interesting fact is that the profiles in Korea and Japan differ radically from all other East Asian countries; their profiles are very close to profiles in Western countries. Here the “why” question is not tackled. However, it raises the questions of educational systems producing *different* (note above the word *better*) qualities in students’ thinking in the same types of cultures. The issue is also discussed in the final section.

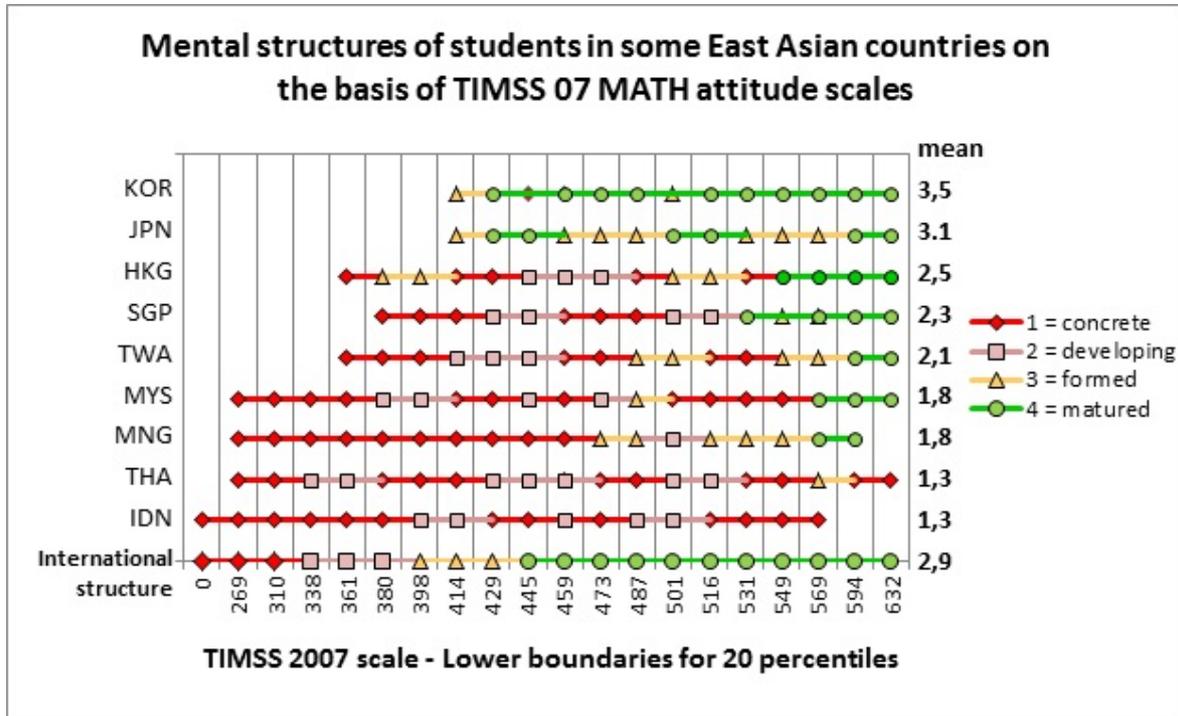


Figure 4. Mental structures of students in some East Asian countries on the basis of TIMSS 2007 MATH attitude scales

3.5 Mental Structures in Different Sets of Countries

Let us calculate the mean rank for each TIMSS country on the basis of country profiles of level of thinking in 20 percentiles. In each percentile, a rank is given and the mean of those ranks in the country indicate the average level of thinking among the eighth graders. This mean rank, ranging 1–4, and the mean achievement score in MATH form a two-dimensional map where all countries are plotted (Figure 5, compare Figure 1 where all the students in each country create one common structure in the country). Heuristically, the countries can be divided into four cultural areas: “Western” structure, “African–Asian” structure, “Middle Eastern–Asian” structure, and “high performing East Asian” structure, all of which are discussed deeper in what follows.

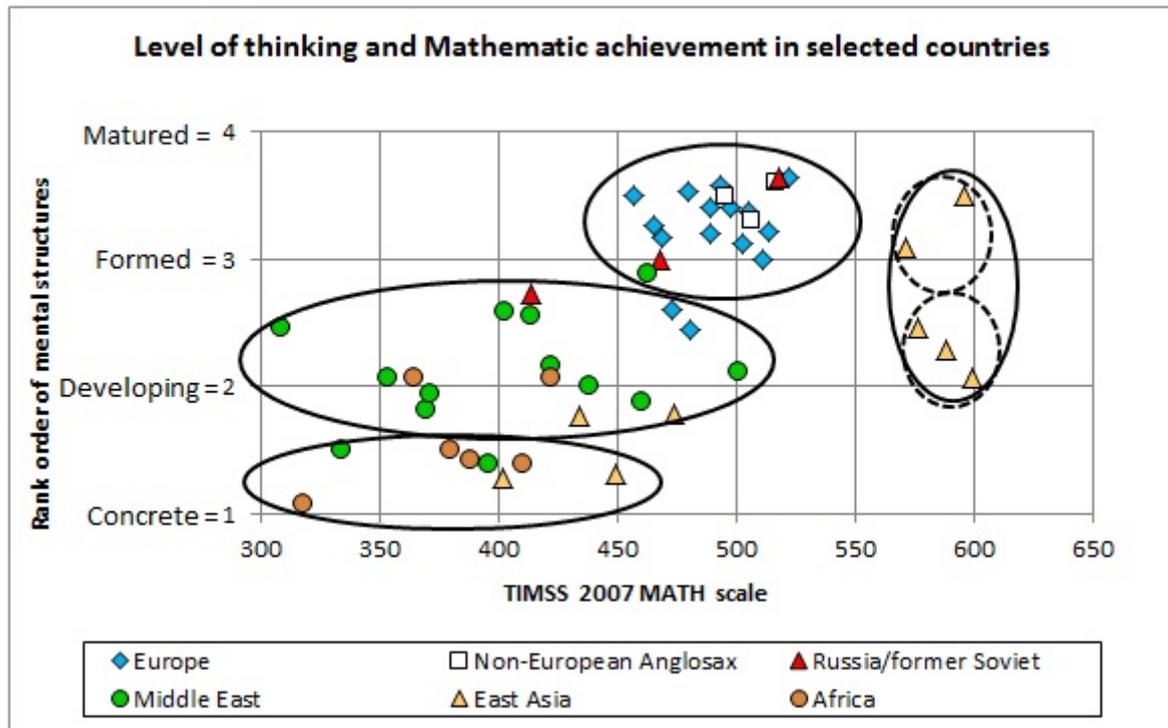


Figure 5. Level of abstract thinking in selected countries on the basis of TIMSS scales for MATH attitude and MATH achievement

All the European countries in TIMSS (excluding Bulgaria and Romania) and all non-European Anglo-Saxon (and Francophone) countries (the USA, Canada, and Australia) as well as Russia, Ukraine, and Israel form a group of countries with a shared structure characterized by (1) higher mean rank than the international average (> 2.9), (2) expected factor structure (see Figure. 1), and (3) high or moderate achievement level. This structure could be called “Western” or “European-North American” structure – “Western” is used because of shortness.

Quite the opposite of the Western structure is the structure in some African countries such as Dzad, Ghana, Morocco, and Egypt that participated in TIMSS 2007, some Middle East countries, such as Saudi Arabia and Syria, and some Asian countries, such as Indonesia and Thailand. Let us call this profile “African-Asian” structure because most of the African countries in TIMSS 2007 share this profile as well as some quite well performing East Asian countries. In these countries, the shared structure is characterized by (1) unified quite much lower mean rank (< 2.1) than the international average showing mainly a “Concrete” structure of thinking, (2) “Totally unstructured” factor structure (see Fig. 1), and (3) low or moderate achievement level. It is possible that the reasons behind the “Totally unstructured” factor structures differ culture-wise. From this perspective, interesting counterparts of a kind are European Bosnia-Herzegovina (BIH, Math score 457) and East Asian Thailand (THA, 449), where the achievement level is somewhat the same, but where the mean ranks of mental structure differ radically from 3.5 (BIH), showing one of the most “Matured” student population, to 1.3 (THA) showing one of the most “Concrete” student population. This fact alone supports the idea that there has to be some cultural factors explaining the phenomenon of answering a certain way in attitude tests. In order to elaborately address with the question it requires quite a deep understanding about the cultures; the issue is not tackled in-depth in this article.

The third group of countries include most of the Middle Eastern countries in TIMSS 2007, some East Asian countries (Mongolia and Malaysia—see also later group four), some African countries (Botswana and Tunisia), and Georgia in Eastern Europe. Let us call this profile as “Middle Eastern-Asian” structure. This group of countries differ radically from each other when it comes to the achievement level, varying from 365 (Botswana) to 501 (Armenia). However, the mean rank in most cases draws near 3 showing that (1) the mental structure is mainly “Developing” or in some cases, is near the international mean (2.9) showing “Formed” but not “Matured” mental structure, and (2) in most cases the general factor structure is either “totally unstructured” or “moderately unstructured” (compare Fig. 1). An interesting block of countries is Bulgaria and Romania, where

the factor structure differs quite radically from those of the other European countries; Bulgarian students' factor structure is "totally unstructured" and the Romanian students' factor structure is "moderately unstructured" (see Fig. 1). From this perspective, the mental structure of Bulgarian students is more close to, for example, that of the Armenian students' than their closer neighbors' in Slovenia or in Bosnia–Herzegovina.

The fourth group of countries is interesting because it includes only East Asian countries with very high performance in TIMSS 2007 mathematics test: Singapore, Taiwan, Korea, Hong Kong, and Japan. The pattern in these countries is called "high performing East Asian" structure. It may be possible to split the structures into two: "High performing East Asian – Western structure" (including Japan and Korea) and "High performing East Asian – East Asian structure" (including Taiwan, Hong Kong and Singapore). Except Hong Kong and Taiwan, they share a so-called "moderately unstructured" factor structure (see Figure 1). However, these countries do not share a common structure from the "maturation" of the abstract thinking viewpoint as did the other three groups of countries. This phenomenon can be seen in Figs. 4 and 5: Korea and Japan mainly share the Western structure, where Singapore, Taiwan, and Hong Kong share the Middle Eastern–Asian structure. And still, their performance is far higher than any other countries—and therefore the achievement level does not explain the differences in the mental structure in these countries. The reasons behind the phenomenon are not discussed in this article. However, *these high performing East Asian countries may be the key to fully understand the different mentalities behind different logics to respond in international attitude tests.*

4. Discussion

This article has concentrated on comparing the mental structure in different countries on the basis of TIMSS dataset scales in MATH attitude and MATH achievement. First, it was noted that in several studies the researchers had noted remarkable differences between countries when it comes to the association of attitudes toward mathematics and mathematics achievement. Second, it was noted that the factor structures in different countries are not identical in TIMSS dataset, but they can be divided into "Expected", "Moderately unstructured", "Totally unstructured", and "Unique" structures. Certain kind of shared cultural background of the countries with "Totally unstructured" factor structure—many Islamic countries seem to share this structure—suggested using the term "*mindscape*" to explain the phenomenon. However, a second glance at the phenomenon evidently shows that students' achievement level explains the discrepancy between the expected and unexpected structure quite plausibly as was suggested also, for example, Kadijevich (2006; 2008) on the basis of earlier TIMSS datasets and Metsämuuronen (2012) with this same dataset.

Because of the same dataset and the same methodology, this study carries the same limitations as in Metsämuuronen (2012). Though the results are based on large data, they carry two weaknesses. First, the data includes oversampling in some countries. For example, compared with the sample sizes in Taiwan (N = 4, 046) and Hong Kong (N = 3, 470) in the original set of data there are actually *three* samples of Canada (British Columbia, Ontario, and Quebec, N = 11, 660) and *three* samples of the USA (the USA, Massachusetts, and Minnesota, N = 11, 051) which produces strict over-sampling in dataset for some countries. This evidently has an effect on the results. Second, it is worth noting that when dividing the original dataset into 20 percentiles, none on the percentile shows Normal distribution; in percentiles 2–19 the population is merely uniform than normally distributed and the range in 1st and 20th percentile is much wider than with other groups because of representing the tail populations. There may thus be some estimation error in the parameters of EFA. However, because of the large sample sizes, the results can be taken stable. Naturally, the results are restricted to 8th graders because of the dataset.

On the basis of the huge data (N > 240,000 students) the higher the achievement level, the more "Matured" mental structure the students have, and parallelly, the lower the achievement level is, the more "Concrete" the mental structure. The international data suggests that, on the basis of factor loadings, bivariate correlations between the negative and positive items, and reliabilities of the scales, the mental structures can be divided into four categories: Concrete, Developing, Formed, and Matured mental structures. When giving the phases a numeric rank order from 1 (Concrete) to 4 (Matured), the international mean rank of 20 percentiles is 2.9 indicating that on average the students draw near the level 3, that is, they are at the "Formed" level of thinking. On the basis of 20 percentiles, 55% of the students are at the "Matured level", 15% at the "Formed" level, 15% at the "Developing" level, and as many as 15% of the students are at the "Concrete" level of thinking.

A deeper profiling and comparison of Western and East Asian countries suggest that there is more than just the achievement level which explains the fragmentation of the original factor structure. The country-wise analysis of Western countries and Asian countries evidently show that in some Asian countries the achievement level has nothing to do with the "Totally unstructured" factor structure or division to "Concrete", "Developing", "Formed",

and “Matured” mental structures. Or maybe it is wiser to say that there may be *inseparable linkage* of achievement level and cultural effects in explaining the phenomenon. For example, in Taiwan, Thailand, and Indonesia also those students in the highest quartile show a “concrete” level of thinking. Because of their capacity of solving complicated and abstract mathematical problems, it is evident that the label ‘Concrete’ is not appropriate with these students when talking about mental structure.

The results raise several questions as seed for further development. First, the *results challenge the whole idea of using one common test all over the world to test mental structures*. It especially challenges the idea of using an excessive number of negative items in the international attitude tests. There does not seem to be any problem when testing and comparing students from the same types of cultures—like in students in the United States and in Europe. However, caution is wisdom when it comes to the comparison and inferring something over the datasets of different cultural areas. From the contemporary psychometrical viewpoint, the negative items are important to ensure the consistency of the respondent. However, if the respondent does not understand the abstract meaning in the question (the hypothesis of achievement level affecting the result), or there is no such concept as thinking negatively or express negative issues (the hypothesis of cultural matters affecting the results), the score does not mean anything. As Metsämuuronen (2012) noted, when the reliability of the score is lower than $\alpha = 0.60$ (as it is in the lowest achievement groups), one could use whatsoever internet-based test to achieve the same accuracy—if not even more reliable a result—than by using well-tested and well-documented Fennema–Sherman test. In any case, one should be very careful when interpreting the correlations between attitude scales and achievement scales in international comparisons; the more in-depth analysis of lower group connections should be done. The same challenge is in the higher end of the scale: when the variance in certain high-performing countries is reduced, the correlations may be reduced. Thus, some of the confusing results (e.g., Ma & Xu, 2004) may partly be explained by technical reasons. There may be two different solutions to solve the issue of international attitude testing with a common test. One solution is to reduce the number of negative items as low as possible—in practice to one negative item per dimension. Metsämuuronen (2012) suggests replacing two suspicious items when using the Fennema–Sherman test in the international settings. If doing so, it would be recommendable to start the test construction with pretesting in East Asian or Middle Eastern countries instead of Western countries. Another option is to create a new, international attitude test. Also this procedure should start in East Asian or Middle Eastern countries instead of Western countries. It may actually be quite an interesting exercise to check how well the Western students would manage to succeed in a test which is fitting well the Asian students.

Second, a small interesting detail in the data is that *it is possible to quite accurately predict the rough achievement level of the students only by knowing the correlation between two attitude variables*; the correlation of variables "MATHS IS NOT ONE OF MY STRENGTHS" and "I USUALLY DO WELL IN MATHS" correlates with the achievement group mean almost perfectly ($r = -0.98$). The fact informs mainly the strict connection of achievement level and mental capability. The analysis was not deepened in this area though it may be worthwhile to compare the countries also from this perspective. One can also ask what kind of simple indicators could be created to indicate the mental capacity or abstract thinking. Maybe it could be possible to create simple indicators for “cultural mindscapes”, too.

Third, though not explicitly handling tests for younger children, the *results challenge the idea of using negative items in the attitude tests when testing the lowest grades’ pupils in schools*. It seems evident that the lower the achievement level, the lower the ability to relevantly respond to the negative questions. The hierarchy in the complexity of the statements, as presented earlier, gives a clue as to what kinds of items could be the most suitable for the children with lower abstract ability. Using straight positive questions (Category 1 in Metsämuuronen, 2012) may raise the reliability of the scores. However, the use of negative items (Categories 2 or 3) is justifiable when it comes to higher grades. Figures 3 and 4 hint that there might be country-wise differences on what would be the suitable age to add negative items to the test.

Fourth, as it is evident that the achievement level of students is connected with the attitude scale discrepancy, it is also evident that *there are existing differences in cultural matters when it comes to the educational settings which need not be harmonized*. Suppose all readers share a common understanding that the variability is valuable as itself—there is no need to totally harmonize the thinking, habits, and values to be the same all over the globe. Especially in cases that the educational system can produce good learning outcomes, the intentional harmonizing would most probably destroy something culturally valuable and unique. Another situation is when the achievement level is not high. Then one can justifiably ask whether the reason for low achievement level is caused by the unique national educational decisions or cultural issues. If, for example, in some countries the students were not encouraged to guess the answer when being unsure of the correct alternative it would evidently

affect the results. If in another set of countries the educational system is geared to *remember the things by heart*—that is, to read the textbook as well as possible—but not to encourage students to *use the knowledge in novel situations*, it would evidently affect the results because the international test would not be based on the known textbook. In these cases, there may be a time and place to think whether these cultural issues are so important that the nation can bear the low achievement in the international comparison. In some countries, the linkage of low achievement level and cultural effects may be *inseparable* or even *exacerbated*.

Fifth, it is interesting to note that *in countries with geographically similar cultural backgrounds, the mental characteristics may differ radically from each other*. Bulgaria and Romania are examples of such countries in Europe, which differ from the other European countries. In Africa, these countries are Botswana and Tunisia. In the Middle East, these countries are Iran, Turkey, and Israel. In Asia, these countries are Korea and Japan. This leads to two different questions: on one hand, it can be asked whether some educational systems can produce *better* mental qualities than some other systems, or do the different educational systems produce just *different* qualities in students' thinking. In this case, education, explicated in the national curricula, creates and strengthens the cultural heritage and diversity. On the other hand, perhaps the educational systems have *nothing* to do with the mental structure. In this case, maybe the cultural setting is created during the childhood in the socialization process before the school age. Therefore education, explicated in the national curricula, is unable to change the tacit cultural heritage already built in students' personalities. This issue needs the careful analysis of cultural issues and deep knowledge of educational systems, which is not discussed in detail here.

All the five points may be worth analyzing deeper. Rich TIMSS data connected with deep knowledge of education, anthropology, psychology, sociology, as well as comparative religion may enable these kinds of comparisons. These kinds of contributions would enrich the research path opened by Metsämuuronen (2012), Bouhlila (2011) and Liou (2010).

References

- Alkhateeb, H. M. (2004). Internal consistency reliability and construct validity of an Arabic translation of the shortened form of the Fennema-Sherman mathematics attitudes scales. *Psychological Reports, 94*(2), 565-571. <http://dx.doi.org/10.2466/pr0.94.2.565-571>
- Areepattamannil, S. (2012). Science Self-Beliefs and Science Achievement of Adolescents in Gulf Cooperation Council Countries. *Educational Studies, 38*(1), 13-17. <http://dx.doi.org/10.1080/03055698.2011.567058>
- Bouhlila, D. S. (2011). The Quality of Secondary Education in the Middle East and North Africa: What Can We Learn from TIMSS' Results? *Compare: A Journal of Comparative and International Education, 41*(3), 327-352. <http://dx.doi.org/10.1080/03057925.2010.539887>
- Caley, M. T., & Sawada, D. (1994). *Mindscapes: The Epistemologies of Magoroh Maruyama*. NY: Gordon & Breach,
- Cronbach, L. J. (1951) Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. <http://dx.doi.org/10.1007/BF02310555>
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Hols, Rinehart & Winston.
- Eklöf, H. (2007). Self-Concept and Valuing of Mathematics in TIMSS 2003: Scale Structure and Relation to Performance in a Swedish Setting. *Scandinavian Journal of Educational Research, 51*(3), 297-313. <http://dx.doi.org/10.1080/00313830701356141>
- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103-127. <http://dx.doi.org/10.1037/a0018053>
- Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman mathematics attitude scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal for Research in Mathematics Education, 7*(5), 324-326.
- Foy, P., & Olson, J. F. (Eds.). (2009). *TIMSS 2007 International Database and User Guide*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <http://timss.bc.edu/timss2007/idb Ug.html>
- Fritz, W. (2006). *Intelligent Systems and their Societies*. E-book online in <http://www.intelligent-systems.com.ar/intsys/index.htm> and <http://www.intelligent-systems.com.ar/intsys/mindscap.htm>

- Gadalla, T. (1999). A Comparison of the Factor Structure of Boys' and Girls' Responses to the TIMSS Mathematics Attitude Questionnaire. Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19–23, 1999). Retrieved from <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED432595>
- Gulliksen, H. (1987/1950). *Theory of Mental Tests*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Hammouri, H. (2004). Attitudinal and Motivational Variables Related to Mathematics Achievement in Jordan: Findings from the Third International Mathematics and Science Study (TIMSS). *Educational Research*, 46(3), 241-257. <http://dx.doi.org/10.1080/0013188042000277313>
- Hembree, R. (1992). Experiments and relational studies in problem solving: A meta-analysis. *Journal for Research in Mathematics Education*, 23(3), 242-273. <http://dx.doi.org/10.2307/749120>
- Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement*, 60(3), 523-531. <http://dx.doi.org/10.1177/00131640021970691>
- House, J. D. (2006a). Cognitive-Motivational Characteristics and Geometry Knowledge of Adolescent Students in Japan: Results from the TIMSS 1999 Assessment. *International Journal of Instructional Media*, 33(1), 95.
- House, J. D. (2006b). Mathematics Beliefs, Instructional Strategies, and Algebra Achievement of Adolescent Students in Japan: Results from the TIMSS 1999 Assessment. *International Journal of Instructional Media*, 33(4), 443-462.
- House, J. D., & Telese, J. A. (2008). Relationships between Student and Instructional Factors and Algebra Achievement of Students in the United States and Japan: An Analysis of TIMSS 2003 Data. *Educational Research and Evaluation*, 14(1), 101-112. <http://dx.doi.org/10.1080/13803610801896679>
- Johansone I., & Malak B. (2008). Translation and National Adaptations of the TIMSS 2007 Assessment and Questionnaires. In J.F. Olson, M.O. Martin & I.V.S. Mullis (Eds.), *TIMSS 2007 Technical Report*. (pp. 63-75). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Kadijevich, Dj. (2006). Developing trustworthy TIMSS background measures: A case study on mathematics attitude. *The Teaching of Mathematics*, 9(2), 41-51. Retrieved from <http://elib.mi.sanu.ac.yu/journals/tm/17/tm924.pdf>
- Kadijevich, Dj. (2008). TIMSS 2003: Relating Dimensions of Mathematics Attitude to Mathematics Achievement. *Zbornik Instituta za pedagogika i strazivanja*, 40(2), 327-346. <http://dx.doi.org/10.2298/ZIPI0802327K>
- Knapp, T. R., & Brown, J. K. (1995). Ten Measurement Commandments That Often Should Be Broken. *Research in Nursing & Health*, 18(5), 465-469. <http://dx.doi.org/10.1002/nur.4770180511>
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151-160. <http://dx.doi.org/10.1007/BF02288391>
- Leung, F. K. S. (2002). Behind the High Achievement of East Asian Students. *Educational Research and Evaluation: An International Journal on Theory and Practice Journal Citation*, 8(1), 87-108. <http://dx.doi.org/10.1076/edre.8.1.87.6920>
- Liou, P.-Y. (2010). Cross-National Comparisons of the Association between Student Motivation for Learning Mathematics and Achievement Linked with School Contexts: Results from TIMSS 2007. ProQuest LLC, Ph.D. Dissertation, University of Minnesota
- Lokan, J., & Greenwood, L. (2000). Mathematics Achievement at Lower Secondary Level in Australia. *Studies in Educational Evaluation*, 26(1), 9-26. [http://dx.doi.org.libproxy.helsinki.fi/10.1016/S0191-491X\(00\)00003-1](http://dx.doi.org.libproxy.helsinki.fi/10.1016/S0191-491X(00)00003-1)
- Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, Mass.: Addison – Wesley Publishing Company.
- Ma, X., & Kishor, N. (1997a). Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. *Journal for Research in Mathematics Education*, 28(1), 26-47.
- Ma, X., & Kishor, N. (1997b). Attitude toward self, social factors, and achievement in mathematics: A meta-analytic review. *Educational Psychology Review*, 9(2), 89-120.

- <http://dx.doi.org/10.1023/A:1024785812050>
- Ma, X., & Xu, J. (2004). Determining the causal ordering between attitude toward mathematics and achievement in mathematics. *American Journal of Education*, 110(3), 256-281. <http://dx.doi.org/10.1016/j.adolescence.2003.11.003>
- Maruyama, M. (1980). Mindscales and Science Theories. *Current Anthropology*, 21(5), 589-599.
- Mehrens, W. A., & Lehman, I. J. (1991). *Measurement and evaluation in education and psychology* (4th ed). Fort Worth: Harcourt Brace College Publishers.
- Melancon, J. G., Thompson, B., & Becnel, S. (1994). Measurement integrity of scores from the Fennema-Sherman mathematics attitudes scales: The attitudes of public school teachers. *Educational and Psychological Measurement*, 54(1), 187-192. <http://dx.doi.org/10.1177/0013164494054001024>
- Metsämuuronen, J. (2012). Challenges of the Fennema-Sherman Test in the International Comparisons. *International Journal of Psychological Studies*. [in process]
- Mulhern, F., & Rae, G. (1998). Development of a shortened form of the Fennema-Sherman mathematics attitudes scales. *Educational and Psychological Measurement*, 58(2), 295-306. <http://dx.doi.org/10.1177/0013164498058002012>
- Mullis, I. V. S. et al. (2008). *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grade*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <http://pirls.bc.edu/timss2007/mathreport.html>.
- OECD (2009). PISA 2006 Technical Report. Retrieved from www.oecd.org/dataoecd/0/47/42025182.pdf
- OECD (2012). PISA 2009 Technical Report. Retrieved from http://www.oecd.org/document/19/0,3746,en_2649_35845621_48577747_1_1_1_1,00.html.
- Olson, J. F., Martin, M. O., & Mullis, I. V. S. (Eds.) (2008). *TIMSS 2007 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <http://pirls.bc.edu/timss2007/mathreport.html>
- Papanastasiou, C. (2000a). Internal and External Factors Affecting Achievement in Mathematics: Some Findings from TIMSS. *Studies in Educational Evaluation*, 26(1), 1-7. [http://dx.doi.org.libproxy.helsinki.fi/10.1016/S0191-491X\(00\)00002-X](http://dx.doi.org.libproxy.helsinki.fi/10.1016/S0191-491X(00)00002-X)
- Papanastasiou, C. (2000b). Effects of Attitudes and Beliefs on Mathematics Achievement. *Studies in Educational Evaluation*, 26(1), 27-42. [http://dx.doi.org.libproxy.helsinki.fi/10.1016/S0191-491X\(00\)00004-3](http://dx.doi.org.libproxy.helsinki.fi/10.1016/S0191-491X(00)00004-3)
- Papanastasiou, E. (2002). Factors That Differentiate Mathematics Students in Cyprus, Hong Kong, and the USA. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 8(1), 129-146. <http://dx.doi.org/10.1076/edre.8.1.129.6919>
- Piaget, J. (1970). *The principles of genetic epistemology*. London: Routledge & Kegan Paul.
- Shen, C. (2002). Revisiting the relationship between students' achievement and their self-perceptions: A cross-national analysis based on the TIMSS 1999 data. *Assessment in Education: Principles, Policy & Practice*, 9(2), 161-184. <http://dx.doi.org/10.1080/0969594022000001913>
- Shen, C., & Tam, H. P. (2008). The Paradoxical Relationship between Student Achievement and Self-Perception: A Cross-National Analysis Based on Three Waves of TIMSS Data. *Educational Research and Evaluation*, 14(1), 87-100. <http://dx.doi.org/10.1080/13803610801896653>
- Stevenson, H. W. (1998). A Study of Three Cultures: Germany, Japan and the United States-An Overview of the TIMSS Case Study Project. *Phi Delta Kappan*, 79(7), 524-529.
- Tapia, M., & Marsh, II G. E. (2005). An instrument to measure mathematics attitudes. *Academic Exchange Quarterly*, 8(2). Retrieved from <http://www.rapidintellect.com/AEQweb/cho253441.htm>
- Van den Broeck, A., Opdenakker, M.-C., & Van Damme, J. (2005). The Effects of Student Characteristics on Mathematics Achievement in Flemish TIMSS 1999 Data. *Educational Research and Evaluation*, 11(2), 107-121. <http://dx.doi.org/10.1080/13803610500110745>
- Vehkalahti, K. (2000). *Reliability of Measurement Scales*. Statistical Research Reports 17. Finnish Statistical Society. Retrieved from <http://ethesis.helsinki.fi/julkaisut/val/tilas/vk/vehkalahti/>

- Vezeau, C., Chouinard, R., Bouffard, T., & Couture, N. (1998). Adaptation et validation des échelles de Fennema et Sherman sur les attitudes en mathématique des élèves du secondaire. *Canadian Journal of Behavioural Science*, 30(2), 137-140. <http://dx.doi.org/10.1037/h0085804>
- Webster, B. J., & Fisher, D. L. (2000). Accounting for Variation in Science and Mathematics Achievement: A Multilevel Analysis of Australian Data; Third International Mathematics and Science Study (TIMSS). *School Effectiveness and School Improvement*, 11(3), 339-360. [http://dx.doi.org/10.1076/0924-3453\(200009\)11:3;1-G;FT339](http://dx.doi.org/10.1076/0924-3453(200009)11:3;1-G;FT339)
- Wilkins, J. L. M. (2004). Mathematics and science self-concept: An international investigation. *The Journal of Experimental Education*, 72(4), 331-346. <http://dx.doi.org/10.3200/JEXE.72.4.331-346>
- Yoshino, A. (2012). The Relationship between Self-Concept and Achievement in TIMSS 2007: A Comparison between American and Japanese Students. *International Review of Education*, 58(2), 199-219. <http://dx.doi.org/10.1007/s11159-012-9283-7>