Topic Study Group 43

Research and Development in Testing (National and International) in Mathematics Education

Ivan Vysotskiy[1], Julia Tyurina[2], Anastasiia Demchenckova[2]

1. Aims of TSG-43

The group based on topics related to the testing as an important aspect of evaluation and also its links in the reform of the mathematics curriculum at the national and international level. The presentations are studies on several interconnected topics.

1) What are the results of development and research of the testing with mathematics content and mathematics competency, their connection?
2) How effective is the use of the LMT measures as a way of comparing teachers' international mathematical knowledge based on various concepts of the necessary knowledge, and hence views on the possibility of measuring them.
3) How typical characteristics of each country can influence the organization of their educational system.
4) What is the evaluation of mathematics education and how examination of the new mathematical knowledge can guide the teaching behavior development along the scientific and rational directions.
5) What is the role of mathematics in educational policies?

1.1. Organizing team

Representatives from different countries were involved in TSG-43 as chair, co-chair, moderators and speakers: China, Japan, the Russian Federation, United States, Brazil, Slovakia, and Norway, that highlights co-thinking and common views not only national, but also international collaborative community of people acting in mathematical education.

The organizing team of the TSG consists of

Chair: Ivan Vysotskiy (Moscow Centre for Continuous Math Education, Russia)

Co-chair: Fumi Ginshima (National Institute for Educational Policy Research, Japan)

[1] Moscow Centre for Continuous Math Education, Moscow, Russia.
E-mail: i_r_vysotsky@hotmail.com
[2] Moscow Centre for Continuous Math Education, Moscow, Russia.

Members:

Richard T. Houang (Michigan State University, US)

Maria Isabel Ramalho Ortigão (Rio de Janeiro State University, Brazil)

Lidong Wang (Beijing Normal University, China)

1.2. Other participants

Besides these team members, participants also include:

Ekaterina Kuksa (Moscow Centre for Continuous Math Education, Russia)

Tibor Marcinek (Central Michigan University, USA)

Jiangong Dong (Wuhu Institute of Educational Science, China)

Joaquim Pinto (Universidade de Aveiro, Portugal)

Bruno Damien da Costa Paes Jürgensen (State University of Campinas, Brazil)

Yu Fu (Beijing Normal University, China)

2. Sessions and presentations

There were 6 high quality presentations in TSG-43. The ICME organizing committee granted the TSG 2 sessions for presentations, but one of them wasn't presented (Tab. 1). Thus, during the hybrid conference only 5 papers were presented.

Tab. 1. List of the presentations

	Paper and author(s)
[1]	On the eighth-grade mathematics achievement and its effect factors — based on seven areas study. *Chunxia Qi* (China)*, Ruilin Wang* (China)*, Qi Huang* (USA)*, and* ***Yu Fu*** (China).
[2]	International comparisons of teacher knowledge: the case of the LMT measures. ***Tibor Marcinek*** (USA)*, Arne Jakobsen* (Norway)*, and Edita Partová* (Slovakia).
[3]	PISA assessment of Brazilian students' mathematical literacy. ***Maria Isabel Ramalho Ortigão*** (Brazil).
[4]	On composing distractors for multiple choice problems. ***Kuksa Ekaterina*** (wasn't presented) (Russia).
[5]	How Chinese design mathematics test. ***Jiangong Dong*** (China).
[6]	Reflections on large-scale assessment and the formatting power of mathematics. ***Bruno Damien da Costa Paes Jürgensen and*** *Mara Regina Lemes De Sordi* (Brazil)

In the first 90-minute session, the Chair of TSG-43, Ivan Vysotskiy, introduced the rest of the Team, described the format of the session and conveyed greetings to all participants from the organizers of the Congress and PC, called on the participants to the most informal discussion possible. As a rule, each presentation took 20–10 minutes, followed by a 10-minute collective discussion. Throughout the two days, TSG-43 reviewed all the research related to the topic of testing, discussed the problems that arise during the preparation, testing, and discussed possible solutions.

The majority of the work can be summarized across five connected themes. As part of the work, researchers consistently, step by step, revealed the main topic of

research and development in testing (national and international) in mathematics education.

We give a brief overview of the presentations as follows.

2.1. Session 1

First talk[1] given by Fu was on the work collaborated by Qi, Wang and Huang. They tried to design math test and questionnaire for 170748 eighth grade students from seven areas in mainland, China. The results showed an inextricable relationship of two parts of the test — mathematics content, and mathematics competency. The mathematics content includes three areas: number and algebra, shape and space, statistics and probability. The mathematics competencies consist of 4 components: knowing, understanding, grasping and application. Eight-grade students achieved high scores, and their pass rate was 92.4%. However, their level of application was lower than the level of knowing, understanding, and grasping. Among all the four components in questionnaire survey, teaching representation had the greatest impact on students' performance, teaching methods and mathematics representation were the second while thinking tendency had a negative effect on students' performance. They also confirmed that development of tools in test booklet and the influence factor in the questionnaire need to be further explored.

The talk[2] given by Marcinek was on his work collaborated by Jakobsen and Partová. It draws on the body of international research conducted in relation to the measures of the mathematical knowledge for teaching developed as part of the Learning Mathematics for Teaching project at the University of Michigan (the LMT measures). There are various conceptualizations of the knowledge teachers need in the work of teaching mathematics. Which means there are different views on feasibility of measuring such knowledge. There were made comparisons between Slovak and Norwegian primary teachers. Three layers of studying were used — Technical layer, Local layer, Global layer. The similarities helped to compare translation-related issues while differences helped to model the issues that likely exist in education systems around the world. The possibility to design LMT forms with close psychometric properties does not imply the meaningful comparisons of teachers' MKT. Local layer provides the most useful information to researchers and constitutes the most important contribution of LMT adaptation literature. Such comparisons might be especially hard to interpret if countries with differences in curriculum, grade bandings, teacher education and certification are compared.

Ortigão[3] presented the results of an investigation focused on analyzing PISA Mathematics items based on Differential Item Functioning (DIF) analysis, to compare results between Brazil and Portugal. Based on descriptive analysis were differences in cognitive skills between the assessed groups. Typical characteristics of each country can influence the organization of their educational system. Knowing these features based on items that favor certain groups, in addition to perceiving the incidence of patterns, is undoubtedly the great contribution from DIF (the identification of items

that disregard one of the main assumptions of the IRT) analysis to educational assessment.

2.2. Session 2

On the final session, the participants discussed two topics.

First, Dong[5] called for discussion on the paper of the tests as aspect of evaluation and as an important link in the reform of the mathematics curriculum in middle school. The evaluation of mathematics education includes not only curriculum evaluation but also teaching evaluation, and mathematics testing is an important aspect of the evaluation of mathematics education. The function of testing can be divided into two types — first is to summarize the results of the entire mathematics education stage, to identify the teaching effectiveness or achievement at the end of the whole teaching stage, the other is to gain feedback with purpose to improve the teaching process, to understand the problems and defects in teaching process at the end of each teaching unit. The focus of the study on mathematics evaluation question setting in Dong's paper is on the level of senior high school entrance examination. Improving the quality of question setters and making the tests papers play an evaluation function more scientifically and reasonably is the top priority. Dong suggested certain basic principles for test items setting, and emphasized that the teachers creating the tests questions must firstly understand the structure, and then start from every single question. The new mathematical knowledge has enriched and optimized the mathematical thinking method in the middle school mathematics, further expanded the application space of the knowledge, which will be an important source of original test questions. Examination of the new mathematical knowledge and scientific solving of a certain type of problems step by step, may better guide the teaching behavior development along the scientific and rational directions.

The main goal of another research[6] by Jürgensen and De Sordi was to explore and reflect upon the issues concerning largescale assessments in São Paulo, Brazil, and the market-oriented policies underlying them. It was about the role of mathematics in educational policies, since they have a great impact on teachers, their everyday life in schools and, consequently, on society as a whole and the formation of students. During the research they did Questionnaire answered by 26 mathematics teachers and Interviews with 10 of them on this topic. The state began to adopt differentiated payment of teachers and other state education employees, through a salary bonus according to the achievement of goals set by the government for each school. External and standardized assessments impose a teaching and learning method that does not consider the different realities. Standardization of teaching (in authors' opinion) is flawed. The expense or cost of these assessments is too high in relation to their return. The authors believe that the answers, the results obtained, should change this evaluation system, as it has been in the same way for a long time and they have not observed significant changes. There should have an investment in teacher training and better working conditions to thus develop the skills expected for students, and not just

charge in the form of assessment and teacher's penalty through bonuses cancel. More than describing the educational reality, the indicators of quality produced by mathematical models and their subsequent dissemination have altered the work routine of teachers and school teams. Mathematics, in general, goes unnoticed, invisible, in these processes, mainly because it comes "in a package" (Skovsmose and Yasukawa, 2009). Therefore, it is pertinent to open the package and proceed to the questions posed by the authors: "what's in the package?", "Whose package is it?" And "what is done by means of the package?". It is necessary to look beyond technical specifications to reverse this picture and to think of an assessment that addresses the complexity of the educational process (including mathematics) and makes it accessible to the general public. Everyone can and should participate in the definition of what is a "good quality" for education, as long as everyone is involved in this endeavor since its elaboration.

Reference

Ole Skovsmose and Keiko Yasukawa (2009). Formatting Power of "Mathematics in a Package": A Challenge for Social Theorising? In: *Critical Issues in Mathematics Education,* Paul Ernest, Brian Greer, and Bharath Sriraman (eds.). Information Age Publishing.