

Topic Study Group 42

Research and Development in Assessment in Mathematics Education

Abid Sohail¹ and Caroline Long²

ABSTRACT TSG-42 provided a forum to share and discuss research and development in the field of Assessment in Mathematics Education. Many interesting and outstanding questions about the nature of interrelationships among assessment and teaching and learning of mathematics, were asked. Recent research has demonstrated the wide range of theoretical and methodological resources that can contribute to assessment in mathematics, including the use of technology. The papers in TSG-42 included reporting on a particular assessment topic or theme, providing the details of an empirical study, giving an exposition of particular assessment practice, or a reflecting on classroom-based assessment.

Keywords: Assessment; Formative assessment; Summative assessment; Large-Scale assessment, Assessment cycle.

1. Objectives

As teachers, practitioners, academics and researchers it is our prime responsibility to conceptualize, debate and formulate learning and assessment systems that prepare our future generations for opportunities and challenges that they may encounter. Assessment is a wide-ranging, multidimensional and vital process integral to teaching and learning. The purposes of assessment can be summarized as being formative, directed at the improvement of teaching and learning, and summative, where the focus is on evaluation of current proficiency, comparability, or evaluating the functioning of an education system as a whole. Various types and formats of assessment support these purposes. Each type of assessment with a well-defined purpose provides specific and useful information to improve standards and quality of teaching and learning. Also, this specific and useful information is beneficial for research. In classroom-based assessment, the interactive teaching, learning and assessment cycle is managed by the teacher, adhering to the perspective provided by policies, procedures and norms of the institutes or states. The assessment cycle may be specifically formulated for a particular cohort of students. In large-scale assessment, this cycle is somewhat extended and generalized to reflect the perspectives and processes which are applicable across countries rather than specific to a certain context. Large-scale assessments have the

¹Examination Board, Aga Khan University, Karachi, Pakistan. E-mail: abid.sohail@aku.edu

²The University of Johannesburg, Johannesburg, South Africa. E-mail: clong@uj.ac.za

potential to provide comparative information about a country's curriculum and teaching practices generally. The purpose of such assessments, the design and development of instruments and the interpretation of results are factors that affect individual countries and influence internal assessment practices. In the 21st century, we have seen new trends and developments in the field of mathematics assessment, including the assessment of the set of skills that encompass creativity, collaboration, communication and problem-solving. New models have been introduced, many of which have encompassed computer-based testing. Also, in this century, the use of Item Response Theory and Rasch Measurement Theory has influenced the design of tests and the analysis and interpretation of results. For this TSG-42, we invited research based on the recent trends and developments in the field of mathematics assessment which cater to the needs of the 21st century. The research papers, presentations and discussions were such that they were beneficial, communicable and accessible to all the stakeholders, would inform a range of assessment practices, and therefore would contribute to making the teaching and learning of mathematics meaningful. Contributions included studies covering (but not limited to) the following themes:

- Theoretical, philosophical and ethical perspectives and debates concerning the assessment of mathematics proficiency;
- Alternative perspectives, models and practices of assessment;
- Classroom-based assessment (formal or informal assessment);
- Teachers and assessment. What is the role of the teacher in classroom-based assessment? What is the impact of high-stakes assessment? How does the phenomenon of “teaching to the test” play out in various contexts?
- Students and assessment. How do different types of assessment affect student learning and motivation? What is the role of feedback in a learner's life? What is the impact of standardised assessments on learning?

What to test? How is a cognitive focus or cognitive development focus accommodated in a testing programme? How is extended problem solving assessed?

- Test design, construction and administration (theoretical, technical and practical components). How do the underlying assumptions of classical test theory, item response theory and Rasch measurement theory affect the design of testing programmes?
- Technology and computer-based assessment;
- Large scale assessment (perspectives, benefits and limitations);
- Validity and reliability: whether or not a test may report dimensions and types of validity and reliability.

The statement that assessment drives learning was illustrated by many thought-provoking presentations at the conference.

2. Sessions

In TSG-42, there were 45 contributions in total 3 sessions. Sessions 1 lasted 90 minutes. There were 9 short oral presentations of 10 minutes. In the second time slot of 90 minutes, there were 6 long oral presentations of 15 minutes of duration. In the third

time slot of 120 minutes, there were 2 long oral presentations of 15 minutes of duration and 9 short presentations. The details are as follows (Tab. 1).

Tab. 1. List of papers presented

Paper and author(s)	
Session 1	
[1]	Students' difficulties in the management of algebraic expression highlighted in large-scale assessment. Federica Ferretti (Italy).
[2]	In-service teachers marking students' answers containing derivation errors. Alberto Arnal-Bailera, José M. Muñoz-Escolan, and Antonio M. Oller-Marcén (Spain).
[3]	Investigating teachers' awareness of the reasons for students' math errors at primary school level. Valentina Vaccaro and Eleonora Faggiano (Italy).
[4]	Cognitive load reduction in math items: performance, gender and socioeconomic status. Emiliano Augusto Chagas and Mauricio Urban Kleinke (Brazil).
[5]	Expressions of mathematical proficiency in students' mathematical work. Priscila D. Corrêa (Canada).
[6]	Structural features in classroom level standardized mathematics achievement results. Timothy Sibbald (Canada).
[7]	Philosophical insights into PISA and mathematics education policy issues. Ian Cantley (UK).
[8]	A unique item format to assess attentiveness to students' mathematical ideas. Ya Mo, Laurie Cavey, Michele Carney, Tatia Totorica, and Patrick Lowenthal (USA).
[9]	Developing preservice elementary teachers' capacity in the design of authentic mathematics assessment. Kim Koh, Olive Chapman, and Shimeng Liu (Canada).
Session 2	
[10]	Evaluating mathematics teachers' professional learning in a PLN: a complex systems perspective. Xiong Wang (Canada).
[11]	Validity of assessments in mathematical textbooks: a study of beginning of primary school level textbook assessments. Grapin Nadine (France).
[12]	Are the stakes the same? A comparison of three types of large-scale assessments in Alberta, Canada. Richelle Marynowski (Canada).
[13]	Factors related to mathematics teachers pedagogic discretion, specifically when evaluating parabolic sketches. Shai Olsher and Kawthar Nakhsh Khalaila (Israel).
[14]	Assessment based on gamification in Hungarian secondary mathematics classes. Marta Barbarics (Hungary).
[15]	I know all about this mathematical topic, but I cannot answer this question' moment, can I have a clue please? Anne D'Arcy-Warmington (Australia).
Session 3	
[16]	Investigating the treatment of missing data in an Olympiad-type test — the case for selection validity. Caroline Long, Johann Engelbrecht, and Vanessa Scherman (South Africa).
[17]	Mathematics assessment practices of primary school teachers in France. Nathalie Sayac and Michiel Veldhuis (France).
[18]	The role of formative assessment experiences in the teaching and learning of mathematics. Adri van der Nest, Caroline Long, and Johann Engelbrecht (South Africa).
[19]	Assessing math in teacher training; what to learn from our students' research. Willem van der Vegt (The Netherlands).
[20]	Transformative assessment system in mathematics education: engaging mind, body and soul. Basanta Raj Lamichhane (Nepal).
[21]	Analysing students' errors in solving context-based problems in Marwa assessment. Umyy Salmah, Uki Rahmawati and Bungkus Dias Prasetyo (Indonesia).
[22]	Raw scores or rasch measures? Lessons from Rasch analysis of secondary one mathematics test. Hairon Salleh, Foo Kum Fong, and Koh Wei Xun (Singapore).
[23]	Research on the level division of mathematical logical reasoning literacy based on solo taxonomy theory. Hua Wu, Junhan Liu, and Fengqi Zhai (China).
[24]	Quality of mathematical reasoning in a Philippine senior high school's pre-calculus examinations on conic sections. Vitus Paul L. de Jesus (Philippines).
[25]	The Results of large-scale assessment as tools for mathematics activity design. Alessandro Gambini and Roberto Capone (Italy).
[26]	Research on the assessment system combining standardization and non-standardization in the mathematics education of top talents. Jianren Niu, Li Lai, Chaodong Chen, Zhirong He, and Liang Yang (China).

3. Posters

The presenters from different regions of the globe also shared their learning experiences through poster presentations. These posters provided a chance to look at different practices and innovate ideas from/for practitioners and researchers. These were 10 posters. The details of the posters are as follows (Tab. 2).

Tab. 2. List of posters presented

Paper and author(s)
[27] Existing assessment practices: a detrimental factor for the value of cognitive diversity in mathematics classroom. Shiva Datta Dawadi (Nepal).
[28] Vertically equating the PSM3 and PSM4. Jonathan D. Bostic, Gabriel T. Matney, Toni A. Sondergeld, and Gregory Stone (USA).
[29] Contents-specifics in teachers' assessment of non-cognitive skills in mathematics education. Enomoto Satoshi, Iwata Koji, Sasa Hiroyuki, Nakagawa Hiroyuki, and Aoyama Kazuhiro (Japan).
[30] A case study of the assessment process in Japanese math classes. Shigeki Kitajima (Japan).
[31] Impact of the standardized test in the classroom: a proposal from the socio-epistemological theory of educational mathematics. Beatriz Elena Martínez Díaz and Ricardo Arnoldo Cantoral Uriza (Mexico).
[32] Semi-automated assessment for mathematical proficiency: the ultimate time-saver for extensive feedback and reliable grades? Filip Moons and Ellen Vandervieren (Belgium).
[33] Standardized testing administration time differences on problem-solving outcomes. Toni A. Sondergeld, Gregory E. Stone, Jonathan Bostic, and Gabriel Matney (USA).
[34] Development of mathematics items with dynamic objects for computer-based assessment using tablet PC. Fumiko Yasuno, Keiichi Nishimura, Seiya Negami, and Yukihiko Namikawa, Jin-ichi Itoh (Japan).
[35] Making classroom assessment happen in novice teachers' class through assessment techniques design. Xiaoyan Zhao and Lingchun Kong (China).
[36] Training and e-assessment of mathematical courses by Xpress-tutor. Philip Slobodsky, Mariana Durcheva (Israel), and Alexander Ocheretovoy (Russia).

4. Way Forward

The field of research in mathematics education assessment is broad and exciting. Submissions were received from countries across the globe. The range of papers included research focused on large-scale studies and national studies, on both primary school, high school, and tertiary levels, but what was most interesting is the number of papers dealing with alternate type assessment.

Included in the large-scale assessment were philosophical insights into PISA (Northern Ireland)^[7], a comparison of different types of large-scale assessment (Canada)^[12], and an investigation into the handling of missing data (South Africa)^[16]. We regard the scrutiny of large-scale assessments both for what they can reveal to education systems, and for their limitations, as important themes to be taken up at future conferences.

Highlights of the papers focused on primary school included assessment practices specific to the primary school (France)^[11,17], and the validity of the textbook assessments (France)^[11]. Focused on the high school were papers on specific mathematical topics. It was most interesting to listen to papers that pushed the boundaries of conventional assessment, and included a number of factors apart from the

merely academic that impacted on achievement. A study looking at engaging mind, body and soul, emanated from Nepal^[20], a focus on mathematical reasoning emerged from the Philippines^[24], and the focus on context-based problems emerged from Indonesia^[21]. While the focus on standard assessment and the improvement of this type of assessment is needed, the investigation of assessment which is not in the main stream is most important for the future. We envisage more papers that are looking at alternate forms of assessment, and that assess what has not been assessed previously. Gert Biesta, originally from the Netherlands, challenged the education community to not only value what can be measured. In the case of many of the papers in ICME-14, there is the attempt to measure what is valued, though this is not always easy.

At the tertiary level, the design of authentic assessment was presented (Canada)^[9], as well as the assessment of mathematics teachers professional learning (Canada)^[10]. A further study looked at the assessment of mathematics in teacher training (Netherlands)^[19]. Poonam Batra, working in teacher education in India, stated that if we want change in our educational systems, we have to empower the teachers, and enable professional teacher agency. How assessment practices can assist this is a good question.

Overall the tendency for research in mathematics education assessment is to broaden its scope to include innovative and all-encompassing characteristics. This trend we value as we look to the future. On the other hand, the improvement of standard assessments to ensure validity and reliability should remain a focus on this topic specific group.

As for meaningful engagement with teaching and learning, the role of technology is critical. In future, technology will not only be a tool of learning and teaching rather it will become a part of the process. Henceforward, besides looking at other important aspects, types and functions of assessments, the role of technology in assessment needs our attention.

We look forward to welcoming an equally diverse range of papers that both improves existing practices and offer new practices.