

TSG 27: Research and development in assessment and testing in mathematics education

Team Chairs: *Marja van den Heuvel-Panhuizen*, Utrecht University, The Netherlands¹
Thomas Romberg, University of Wisconsin – Madison, USA

Team Members: *Vladimir Burjan*, EXAM, The Slovak Republic
Thabiso Nyabanyaba, University of the Witwatersrand, Johannesburg, South Africa
Yoshinori Shimizu, Tokyo Gakugei University², Japan

Aims and focus

The purpose of TSG 27 was to investigate recent developments in assessment and testing in mathematics education and to provide the participants with a forum for sharing and discussing these developments. To organize the TSG two themes were followed:

- Research and development in external assessments.
- Research and development in classroom assessments.

The way the topic study group was organized gave an unique opportunity to connect two different ways of assessing mathematical understanding which belong to two different assessment worlds with each their own purposes and approaches, but which at a closer look turned out to have more in common than might have been expected. Having the possibility to connect these two assessment worlds worked out to be very enriching to our thinking about future developments in assessment.

External assessment

External assessments are tests designed by a source (e.g. state agency, test publisher, researcher) external to the mathematics classroom, and are administered via a prescribed set of procedures. Such tests are often written but they may be oral as well. There are at least three types of such tests that differ in terms of how the information about student performance is derived from their use, and how the information is used. First, *profile tests*, such as the Third International Mathematics and Science Study (TIMSS), or OECD's Programme for International Student Achievement (PISA), are designed to present policy makers with information about a population or subgroups of a population of students. No summary information about individual students is possible. Second, *research tests* are designed by researchers to gather evidence in order to verify some assertion, test a hypothesis, etc. Administration of such instruments may either be for gathering information about groups of students or about individual students. Finally, *objective tests* are designed so that the information derived can be used to make decisions about individual students. The information may be used for diagnostic purposes, meeting performance criteria, admission to programs, and so forth.

Max Stephens of the University of Melbourne (Australia) presented an overview paper on research on external assessments. This presentation examined four aspects of recent work. First, recent developments in external assessment at the end of high school, including assessments used for university entry have tended to blur the distinction between

¹ Today, MvdHP's affiliation is Utrecht University, The Netherlands, and Humboldt University, Berlin, Germany.

² Today, YS's affiliation is Tsukuba University, Japan.



I C M E
1 0
2 0 0 4

TSG

Topic Study
Group 27



I C M E
1 0
2 0 0 4

TSG

Topic Study
Group 27

internal assessment carried out by teachers and externally mandated assessment. Second, the growing use of technology in assessments is important because of the use of technology by students in externally prepared assessments of mathematics. It is also important as a means of delivering externally constructed assessment tasks to the classroom. This opens up new possibilities for various forms of profile tests that are mandated for use in schools by national or state or local education authorities. Technology delivered assessment challenges us to consider what is assessed and who is ultimately responsible for assessment in mathematics. Third, the impact of profile tests, including those that are mandated by educational authorities, as well as those that are part of external international assessment in mathematics, such as TIMSS and PISA was addressed. Fourth, the growing interface between instruments developed by researchers to assess students' mathematical understanding and the use of these instruments in system-endorsed programs of school improvement and teacher professional development was examined.

Nine papers were presented or made available for participants on this topic. The papers that were presented were by:

- *Dvora Gorev* (Israel). This paper dealt with how students handle a final exam in calculus that is based on tasks using the computer, and how a computer-based environment can improve low achiever's presentation of understanding.
- *John Threlfall and Peter Pool* (UK). This paper described the consequences of creating a computer environment to what is being assessed. Different from paper and pencil tests, computer items have the potential to offer a dynamic and interactive environment in which mathematical thinking may arise and can be assessed.
- *Murad Jurdak* (Lebanon). This paper described the analysis and development of assessment tasks on problem solving by means of action maps (a schematic representation of the organization and sequence of the actions of an activity).
- *Brian Doig* (Australia). This paper dealt with the development of a formal assessment instrument for young children. The items of this instrument were presented orally in a 'lock-step' fashion (that is, all children were working on the same question at the same time, and advanced through the questions at the same pace). The children could answer the questions by ticking a picture or writing a number.

The papers that were made available were by:

- *M. Pedro Huerta, Eduardo Galán, and Ramon Grandell* (Spain). This paper dealt with the possibilities of using concept maps built by the students as an assessment tool.
- *Lázaro S. Dibut Toledo, Narciso R. De León Rodríguez, Eduardo Backhoff Escudero, José Luis Ramírez Cuevas, and Héctor León Velazco* (Cuba and Mexico). This paper described the development of an on-line university entrance exam via collaboration between faculty in two countries.
- *Signe E. Kastberg and Beatriz S. D'Ambrosio* (USA). This paper deals with what understanding students have when they solve mathematical context problems



I C M E
1 0
2 0 0 4

TSG

Topic Study
Group 27

from NAEP assessments in mathematics and other subjects and with the consequences for assessment design.

- *Göta Eriksson* (Sweden). This paper deals with assessing and teaching young special education children. It is argued that we are obliged to view children in need of special support as competent learners and logical human beings and that it is our responsibility to understand the child and never give up in finding the child's competence.
- *R. M. Dimitric* (USA). This paper deals with the development and testing of a diagnostic test for elementary statistics for university students.

These papers dealt with a rich variety of topics. Attention was paid to psychometric aspects related to the development of a standardized or a diagnostic test, the consequences and possibilities of using computer or on-line assessment, and the use of action maps and concept maps to improve the assessment from the perspective of the mathematical content that is being assessed. Looking at the trends that were mentioned in the overview paper, the nine papers gave support to the analysis that was made. The papers also made it clear that there is a blurring distinction between internal assessment carried out by teachers and externally mandated assessment, that there is an increasing use of technology in assessment, and, finally, that there is a growing use of external assessments as levers for both school improvement and the professional development of teachers.

Classroom assessment

Classroom assessments are methods used by teachers (or groups of teachers) to gather and document information about individual student performance. To monitor progress, to grade performance, and to modify instruction teachers use the information derived from such assessments in a variety of ways. Mathematics teachers have traditionally monitored their students' progress by giving quizzes and chapter tests, scoring answers, and periodically summarizing student performance in terms of a letter or a number grade. Although often items have been developed externally (e.g. by the text book authors, researchers and test designers) teachers are free to modify, adapt, or add tasks; to develop scoring rubrics; and to include information from observations of student work, or from interviews. Today, because of the reform initiatives throughout the world, teachers are expected to incorporate information from observations, interviews, project work, etc. in their judgments of student performance.

David Webb of the University of Wisconsin-Madison (USA) presented an overview paper on the research on classroom assessment. Although there has been a heightened interest in research on this aspect of classroom practice, researchers' varied perspectives on what constitutes classroom assessment appears to have left this potentially informative line of research languishing as an ill-defined and misunderstood topic. To move the discussion of classroom assessment forward three fundamental aspects of internal assessment were used to distinguish it from external assessment: (1) the influence of teachers' conceptions and experiences, (2) the norms and routines of school contexts, and (3) the central role of pedagogical decision-making. For each aspect the findings from several related studies that collectively position classroom assessment as an essential research context were examined. Given that internal assessment is contextualized and somewhat dependent on teachers' conceptions, norms and practices, it was important to take note

of these factors when engaging in research, as well as interpreting research findings in this domain. The paper concluded with observations on future directions for research drawing upon intersections between classroom assessment, research on pedagogical decision-making, and reform initiatives in mathematics.

Eight papers were presented or made available for participants on this topic. The papers that were presented were by:

- *Ruhama Even* (Israel). This paper dealt with how teachers make sense of assessment data, and how they can be helped to adopt reform assessment procedures.
- *Marie Hofmannová, Jarmila Novotná, and Renata Pípalová* (The Czech Republic). This paper dealt with teacher-made assessment instruments in two schools where mathematics is taught in a foreign language.
- *Hari P. Koirala and Marsha J. Davis* (USA). This paper described the design of an assessment task for judging pre-service teachers' ability to assess high school students' mathematical understanding.
- *Lisa Björklund* (Sweden). This paper dealt with how year-5 teachers in Sweden assess and describe pupils' performance by means of an external instrument.

The papers that were made available were by:

- *Ilana Lavy and Atara Shriki* (Israel). This paper described how class discussion and portfolios could be used in teacher education to assess pre-service mathematics teachers' professional growth.
- *Nellie Verhoef and Harrie Broekman* (The Netherlands). This paper focused on views of the learning and teaching of geometry as a framework for designing classroom assessment materials by teachers.
- *Rosemary Callingham and Patrick Griffin* (Australia). This paper dealt with establishing the validity of external performance assessment tasks for year-10 students when administered by teachers.
- *Pi-Jen Lin* (Taiwan). This paper described procedures for assisting primary school teachers to design assessment tasks and analyze student responses.

In summary, because there are several fundamental aspects of classroom assessment that distinguish it from external assessment, four additional major issues were addressed in these papers. First, the influence of teachers' conceptions and experiences about how mathematics should be assessed and how those conceptions could be changed was examined for both pre-service and in-service teachers. Second, a particular concern was related to how to improve teachers' ability to hear from students and judge more than just answers, as they are involved in current reform initiatives in mathematics. In particular, the ways in which teachers' judgments are related to "learning lines" or "assessment trajectories" were deemed critical. Third, understanding the influence of assessment within the norms and routines of school contexts and teachers' didactical decision-making is critical. Finally, collaboration between teachers and researchers with multiple perspectives is important, given that classroom assessments are contextualized and somewhat dependent on teachers' conceptions, norms and practices.



I C M E
1 0
2 0 0 4

TSG

Topic Study
Group 27

The study group ended with a discussion of “visions for the future.” Three questions were of particular concern. First, how can the vast amount of research on assessment become a more coherent body of information? Coherence is important because at present there are many “blind spots” such as – how do teachers come to know a student’s understanding.

Second, how can assessments be better linked to instruction than is presently the case? In fact, the suggestions by several authors was that use of learning lines, trajectories, progress maps, and so forth can link assessment to what is worthwhile to teach. If so, it gives the field an opportunity to develop a didactical model for assessment design.

Third, how can external and internal assessments be linked? The suggestions included using more standard(ized) assessments in classrooms, as is being done in Sweden, and more teacher-based assessments in external assessment systems.

This report was written by Marja van den Heuvel-Panhuizen and Thomas Romberg. They are happy to be contacted at m.vandenheuvel@fi.uu.nl and tromberg@wisc.edu, respectively, for further information on the work of this TSG.



I C M E
1 0
2 0 0 4

TSG

Topic Study
Group 27