

## Introduction to the papers of TWG21: Assessment in mathematics education

Paola Iannone<sup>1</sup>, Michal Ayalon<sup>2</sup>, Johannes Beck<sup>3</sup>, Jeremy Hodgen<sup>4</sup> and Francesca Morselli<sup>5</sup>

<sup>1</sup>Mathematics Education Center, Loughborough University, UK; [p.iannone@lboro.ac.uk](mailto:p.iannone@lboro.ac.uk)

<sup>2</sup>Department of Mathematics Education, University of Haifa, Israel, [mayalon@edu.haifa.ac.il](mailto:mayalon@edu.haifa.ac.il)

<sup>3</sup>Department of Mathematics Education, University of Wuerzburg, [beck@dmuw.de](mailto:beck@dmuw.de)

<sup>4</sup>UCL Institute of Education, University College London, UK, [jeremy.hodgen@ucl.ac.uk](mailto:jeremy.hodgen@ucl.ac.uk)

<sup>5</sup>DIMA Mathematics Department, University of Genoa, Italy, [morselli@dim.unige.it](mailto:morselli@dim.unige.it)

*Keywords: Summative assessment, formative assessment, validity, feedback, assessment methods*

### Introduction

TWG21 met for the second time in Utrecht at CERME11 and in this conference we sought to continue the work started at CERME10. The aim of the previous meeting was to ascertain where the interest of our community is when thinking about assessment, and to maintain the focus firmly on mathematics. At CERME11 we discussed 14 papers and 3 posters which helped defining such interest. We noticed again a variety of focal points: from validation of large-scale assessment instruments, to the affordances and drawbacks of online assessment – especially in the university context - to the details of construction of individualised feedback. As in the previous meeting the papers also presented a variety of methodologies: from large quantitative studies to more nuanced qualitative investigations. Among the submissions we also received papers related to students' perspectives and teachers' perspectives on assessment. These themes were not prominent in the past meeting of the group and we welcomed the new perspectives they brought. Finally, we decided to group papers together that indicated the role that mathematics has in the assessment: this is to say papers that focus on the specifics of mathematics, such as assessing proof.

### Thematic clusters

We identified 5 overarching themes that will serve to organise the papers submitted to TWG21. Below, we describe each of these themes in turn.

**Large scale standardised assessment:** There is strong interest in our community for the use and validation of large-scale standardized assessment, both for benchmarking national students' attainment and for mapping common misconceptions across educational stages. Garuti and Martignone present the introduction of the INVALSI standardised tests in Italy and then investigate the possibility to assess argumentation through the multiple-choice questions included in the INVALSI tests. By using Toulmin's model (1958) to analyse the structure of arguments included in the tests the authors find that although standardised multiple-choice arguments can identify only certain aspects of argumentation, this is not to say that the aspects that these tests do identify have no educational value. They conclude by proposing that the Toulmin's model (1958) could be used to guide the construction of such tests. Klegseth, Kaspersen and Solstad use Rash Analysis to determine

whether non-standardised test included in Norwegian textbooks are a valid and reliable indicator of mathematical competence. In order to do so the authors focused on the concept of function. The items tested were divided into 5 competencies categories and the authors found that, with a careful selection of such test items, the set they identified can be a reliable and valid measure of mathematical competences for these students. Drüke-Noe and Siller present a newly developed test with easy tasks which assesses basic mathematical competencies that are essential for vocational training. The test is used in years 8 and 9 in Germany. From both a content-oriented and a process-oriented perspective, preliminary test results reveal a substantial lack of these competencies in all academic tracks, and differential effects were found with respect to year and class that need to be examined beyond curricular analyses. Finally, Lasorsa, Garuti, Desimoni, Papa, Costanzo and Ceravolo also worked from the database generated by the INVALSI test for mathematical proficiency in Italy and investigated the psychometric and qualitative-educational properties of a pool of these questions. They seek to find specific cognitive obstacles that the students need to overcome to reach the desired competence levels.

**Assessment with technology and Computer Aided Assessment (CAA):** As in CERME10 we received papers on the use of technology for assessment, both papers at university level. Sangwin reports the development and evaluation of an online linear algebra examination administered using the system STACK (Sangwin, 2013). The aim of this study is to investigate whether common questions in linear algebra could be assessed automatically, which was the case for most of the questions tested. The implication for the assessment of mathematics are that tools need to be created and tested to assess explanation, justification and reasoning, and that the current questions of paper-pencil test should be reconsidered since they often focus on mechanical processes. Barana and Marchisio report the implementation of automatic assessment methods both in classroom activities and in online homework tasks to enact formative assessment strategies. The aim of such implementation was to foster self-regulation in the students. Their findings illustrate how formative assessment strategies can be implemented into such a blended-learning setting and how students perceived the effectiveness of the experiment.

**Assessment of mathematics: the teachers' perspective:** We received four contributions on this topic, all reporting research in school settings. Horoks, Pilet, Coppé, De Simone and Grugeon-Allys present the design of a large-scale questionnaire aimed at analysing teachers' assessment practices and how teachers use the information they gather to promote students' learning, in the context of algebra. The authors discuss the design of suitable questions to gather information on the diversity of teachers' assessing practices. Kaplan, Kan and Haser present a study that investigated preservice middle school mathematics teachers' (PST) professional development related to formative assessment practices and focus on how preservice teachers' practices changed during the teacher education program. The authors explored PSTs' performance on a task where they were asked to analyse a lesson plan without any formative assessment opportunities built in and found PSTs did not always realise the lesson plan lacked any formative assessment, especially if they had not yet attended to measurement and assessment courses and methods of teaching courses. Moreover, PSTs were keener to suggest assessing strategies based on their role as teachers, and less to insert peer assessment into the plan. Sangari, van den Heuvel-Panhuizen, Veldhuis and Gooya report preliminary results of

a survey of Iranian teachers aimed at investigating to what extent teachers engaged with Descriptive Assessment, an approach for evaluating students' achievement by collecting and documenting evidence regarding their learning and performance. The survey showed that teachers engage with some of the formative assessment tasks, although not as frequently as it would be desirable. Finally, Prendergast, Treacy and O'Meara presented a study concerning teachers' perceptions of the Bonus Point Initiative (BPI) in Ireland. This initiative consists in awarding an extra 25 points in final examination results to those students who choose to study mathematics at higher level. The study adopted a mixed-method design. Findings show that, although teachers recognized the value of the BPI initiative in promoting the study of mathematics and recognizing the effort paid by students in studying maths at higher level, many of them feared that the BPI may decrease the standard of students attending to higher level courses.

**Assessment of mathematics: the students' perspective:** These papers focused on the impact of assessment on students' experiences and attitudes both at school and university levels. Demosthenous, Christou and Deetra Pitta-Pantazi propose a framework for designing different types of assessment tasks to elicit evidence about how students respond to the mathematical ideas presented in learning trajectories. The authors illustrate the application of the framework by drawing on a learning trajectory of fraction division for sixth grade students and discuss how the enactment of the trajectory in classroom could be adjusted to students' current understanding. Häsä, Rämö and Viii Virtanen discuss a study part of an ongoing larger project concerning student self-assessment skills in university courses. The authors developed a method enabling large cohorts of students to assess their own learning outcomes and to give their own course grades with the help of an automatic verification system. The paper explores whether the self-assessed grades correspond to the students' actual skills, and how well the automatic system can detect issues in the self-assessment. Based on an expert's evaluation of the skills of two students, the study concludes that although for large part the model works as intended, there are some cases where neither the self-assessment nor the computer verification seem to be accurate. Bóra and Juhász discuss pair-work testing. Their pilot study consisted of four tests: twice students were tested in pairs and twice students were tested individually. The researchers recorded the marks and overall performance of the two-two sets of each type of tests. Online questionnaires revealed that in general students had a positive attitude towards paired testing: they rated positively the concept and they rated their testing experience positively. In addition, the overall level of paired work was high. The authors point to the need to continue the study with a larger number of students to learn more about how paired testing improve collaboration and mastery of skills.

**Aspects specific to the assessment of mathematics:** Starting from the work in our previous meeting we wanted to emphasise the role of mathematics in the assessment, e.g. the specifics of assessing mathematics. We received four contribution to this topic. Bolondi, Ferretti and Santi use the framework of didactical situations (Duval, 1995) to investigate the persistence of certain mathematical misconception across educational stages (e.g. from school to university mathematics). By analysing the data collected in the database generated by the INVALSI test for mathematical proficiency in Italy the authors found that certain misconceptions concerning the operations between exponentials persist across the school to university transition, for low as well as medium attaining

students. Watkins, Lamm, Kohli and Kimani investigate the relation between characteristics, beliefs, background and perception of students on student attainment in the context of US Community Colleges, a much under-researched area. Using linear regression on a large database collected for the study they are able to individuate some predictors of student achievement and conclude that rich professional development opportunities for teachers have the potential to enhance teachers' effectiveness in this context. Cusi and Telsoni describe the design of individualized teaching/learning paths in the context of formative assessment at university level. The paths are designed to be tailored to the students' responses to online test and provide tailored feedback at each stage of the test. Preliminary results show how students interact with the feedback received and highlight that few students are aware both of the difficulties they have in solving the problems and how to overcome them. The individualized teaching/learning paths have the potential to develop the awareness of strategies to overcome difficulties. Finally, Turiano, Boero and Morselli explore the emergence of a 'culture of theorems' (Boero, 2007) through assessment involving narratives in a 10<sup>th</sup> grade classroom. After the students were taught some Euclidian geometry they were given as part of their assessment the task to write to a friend in another school and explain what a theorem is. Preliminary analysis of the narratives shows instances of students' meta-mathematical knowledge regarding the culture of theorems.

## Conclusions

The papers that were presented in the TWG21 indicated once more the variety of interests in the assessment of mathematics community, both in terms of focus and of methodologies. Again we received papers adopting quantitative methods (such as Rasch analysis) and papers focusing on small in depth case studies. This reflects the variety of approaches in our field and it is normal that this variety emerges on the work of our group. However, while in the past meeting we found that authors were concerned by the lack of a shared language to discuss assessment, this time we welcomed more agreement on basic definitions. The emphasis that we detected on students' and teachers' perspectives also indicate the need for investigating how assessment affects those participating in it, following the observation we made at CERME10 of a lack of studies on this topic. Again, we sought to maintain the focus on mathematics – e.g. we highlighted the impact of the discipline characteristics on the way in which we think about assessment, both summative and formative. Finally, we plan to meet again at CERME12 in Bolzano (Italy) to push forward the work of the group and out thinking on the assessment of mathematics.

## References

- Boero, P. (Ed.) (2007). *Theorems in school*. Rotterdam: Sense Publishers.
- Duval, R. (1995). *Sémiosis et pensée humaine: Registres sémiotiques et apprentissages intellectuels*. Berne: Peter Lang.
- Sangwin, C. J. (2013). *Computer aided assessment of mathematics*. Oxford, UK: Oxford University Press.
- Toulmin, S. (1958). *The use of arguments*. Cambridge: University Press.