An introduction to TWG21: Assessment in mathematics education

Francesca Morselli, Christina Drüke-Noe, Chiara Giberti, Gözde Kaplan-Can, Johanna Rämö

An introduction to TWG21: Assessment in mathematics education

[1]Francesca Morselli, [2]Christina Drüke-Noe, [3]Chiara Giberti, [4]Gözde Kaplan-Can,

[5]Johanna Rämö

[1]University of Genova, Italy morselli@dima.unige.it

[2]Pädagogische Hochschule Weingarten, University of Education, Weingarten, Germany

[3]University of Bergamo, Italy

[4]Middle East Technical University, Ankara, Turkey

[5]University of Eastern Finland, Finland

Introduction

This Topic Working Group (TWG21), established in 2017 (CERME 10) in Dublin, met in Budapest for the fifth time. As in the preceding meetings, the contributions to this TWG addressed assessment on many levels and in many contexts of mathematics education, adopting a variety of theoretical approaches and methodological frameworks. The contributions (20 papers and 3 posters) were organized around five themes. For each theme we present a short synthesis of the contributions. Afterwards, we reflect on emerging issues and directions for further research.

The themes

Design and implementation of digital tools for assessment: contributions concerning this theme explore how digital tools can be used to implement formative or summative assessment. Barana, Marchisio, Roman and Sacchet present an example of automatic formative assessment realised within a digital learning environment. They focus on example eliciting tasks, showing the information on the students' solving process that is provided by the environment. A poster by Eumann and Barzel concerns a digital diagnostic tool. They investigate how the tool enables and encourages teachers to use formative assessment.

National and international standardised assessment: from different perspectives, the four contributions belonging to this theme draw attention to the relation between a curriculum, teaching and students' performance in national tests. With respect to a reformed curriculum in Norway, Eriksen and Vos discuss the relation between an intended national curriculum and officially published illustrating tasks that inform about national standardised examinations. Their study reveals that characteristic competences, such as reasoning, exploring, representing and modelling, are more prevalent in higher stands of education than in lower ones. Rewer and Greefrath investigate how much academic language and mathematical difficulty-determining characteristics of test items affect students' performance in a German nationwide standardised test for eighth grade students. They find that the level of cognitive demand of the tasks has a significant influence on students' performance. With respect to linguistic characteristics they show that only the amount of text determines students' performance in the test. Ableitinger, Dorner and Krammer discuss the construction of a valid tool designed to assess procedural knowledge. They also investigate the relationship between success in completing curriculum-related procedural tasks and some student characteristics and their grades in

mathematics. The authors find no disparity between genders in terms of procedural knowledge but find positive statistical correlations between procedural knowledge and general mathematical ability, self-efficacy and a dynamic view of mathematics. As regards mathematical knowledge, Kılıç, Doğan and Kılıç highlight the gap between curricular norms and expectations on the one hand and test scores obtained in national tests on the other. With reference to Turkey, the authors question if an enacted curriculum sufficiently supports students to meet the demands of a High School Entrance Exam.

Students' role in assessment: contributions concerning this theme explore the design and implementation of forms of assessment in which students are the main agents, as well as the way students perceive assessment and the possible effect of assessment on their learning. Rämö, Häsä and Yan discuss an undergraduate mathematics course where students gradually developed self-assessment skills and, finally, were allowed to choose their own course assessment method. During the course, students showed an increasing level of self-efficacy and attitude towards self-assessment. Self-assessment practices also improved during the course, since students relied more and more on internal rather than external feedback. Such results suggest the importance of designing and implementing alternative forms of assessment that promote student agency. Morselli and Quartara address the implementation of formative assessment activities with tenth graders. They present and discuss the design of a sequence of formative assessment activities with technology support. Although the analysis of a collection of individual stories is in progress, they report the findings related to the efficacy of formative assessment strategies in terms of self-regulation by exemplifying one case. The theme of the students' role in assessment also includes issues related to inclusion, diversity and social justice in assessment. The study by Barbero, Cascella, Giberti and Viale investigates the learning gap between foreign and native speakers in Italian national standardised assessment. They build on their previous research which has proposed criteria explaining the disadvantage of foreign students, and investigate whether these criteria might be validated on a new subset of data. The validation is based on two independent analyses of the same test: a quantitative analysis to highlight items with a citizenship DIF and a qualitative analysis to classify items according to the criteria. Their results show that the criteria proposed are useful to interpret the citizenship DIF in almost all items and suggest some possible adjustment of one of the criteria. In their poster, Kmecikova, Ratvajová and Hubeňáková present and study a task for assessing students' ability to order systematically. Their presentation is part of a larger study that aims to support students with language barriers.

Formative and summative assessment classroom practices: contributions to this theme address the design and implementation of classroom assessment practices. Tscholl, Stampfer and Hell present the design of a diagnostic assessment for first year university students and discuss its correlation with academic success, showing that this kind of assessment may be used to identify students at risk of dropping out. The authors suggest the need for further research to use the information provided by diagnostic assessment, in terms of individual (automated) feedback or compulsory attendance to a bridging mathematics course. In an interview study with German secondary school teachers, Drüke-Noe investigates principles which mathematics teachers actively apply when designing class tests, which are the predominant form of internal written assessment. The study implies that the teachers tend to overestimate the coverage of competencies and the cognitive demand of their class tests. Though the design principles of individual teachers differ, the results consistently and clearly show

the influence of exit exams on the design. Højgaard discusses the challenges and affordances of assessing students' mathematical competencies. The study describes a case of using oral exams in competency oriented mathematics education and demonstrates how it has the potential of raising the validity of assessment. Sommerlade and Eichler develop a test with the purpose of assessing conceptual and procedural calculus knowledge. The test is assigned to upper secondary school students and students in the transition from school to university. In their contribution the authors report the results of a preliminary version of the test, identifying scales showing a good reliability and suggesting possible adjustments for scales with low reliability. Lemmo suggests strengthening the formative use of written internal tests. She implies that this approach calls for sharing the data collected by means of these tests to provide both teachers and students with evidence that can be used to support teaching and learning. Linker and Selter present a part of a project that investigates how teachers can be supported in implementing support-oriented diagnostics in inclusive early mathematics teaching. In their poster, they focus on teachers' self-reported practices on formative assessment before and after participating in an in-service program. Data from the questionnaire shows an increase in the overall group's mean from pre to post test.

Teacher education concerning assessment: several contributions investigate the issue of assessment both from the point of view of in-service and preservice teachers education. Zender and Geisen investigate preservice teachers' perspectives toward formative assessment in their university education and school. They also point out how preservice teachers deal with individual feedback. Findings demonstrate that preservice teachers have a positive disposition toward formative assessment both in their university education and in school. Participants also appreciate the opportunity for self-reflection they experience by individual assessment. Moons' Belgian study examines the reliability of grading. He compares different modes of grading. In one mode, assessors know about their gradings' influence on a final grade, in the other one they don't. Moons shows a positive influence of so-called blind grading on the inter-rater reliability. Koklu and Arıcan present the design of a diagnostic assessment for preservice teachers with a focus on conditional probability and Bayes' rule. Results suggest that a robust understanding of conditional probability is not sufficient to master Bayes' rule. Also in this study, results pave the way to a reflection on the way to formatively exploit the results of diagnostic assessment, so as to improve preservice teachers' understanding of the topic. The contribution presented by Kaplan-Can and Haser focuses on preservice middle school mathematics teachers in Turkey. They investigate planned and enacted formative assessment strategies in a task-solving classroom environment. Results highlight that the teacher involved in the study used formative assessment strategies (e.g. peer- and self-assessment strategies) but unintentionally and superficially. Therefore, the importance of offering to preservice teachers practice-based courses built on the theoretical background emerges. Kröger and Büchter consider the role of competences in assessment and, in particular, in written teacher-made tests in Germany, Sweden and Finland. They propose a framework for classifying mathematical competence requirements and they introduce, as a part of the framework, the notion of competency-related tools. The framework and its terminology have an impact both in the field of research and school practice contributing to the general discussion about mathematical competence. The importance of teachers' beliefs about assessment is highlighted in the work proposed by Amado and Morselli, which is based

on the comparison of Portuguese and Italian teachers' answers to the same questionnaire. Their research shows that the assessment features often mentioned by teachers are clarity, strictness and objectivity. Furthermore the authors identify six categories concerning critical aspects of assessment and observe that teachers refer almost exclusively to summative assessment rather than formative assessment. Teachers' perspective on assessment emerges also in the poster presented by Kohanová: the analyses of Norwegian teachers' interviews sheds a light on the challenges faced by teachers as they shift their practice from grades to "gradeless".

Conclusions and directions for further research

Connections and comparisons among the contributions paved the way for a reflection on what is specific to *mathematics* assessment. The TWG21 participants stressed that mathematics assessment should take into account specific features of this subject's teaching and learning processes (e.g. the importance of representations, the linguistic aspects of mathematics, the specific role of mistakes, etc.) as well as results from mathematics education research. Accordingly, despite the variety of themes and methods represented in the contributions, discussions during all the TWG21 sessions evidenced the existence of a common basis of theoretical references, concerning formative assessment practices (Wiliam & Thompson, 2007) as well the notion of competence (Niss & Højgaard, 2019) and the dichotomy between procedural and conceptual understanding (Hiebert & Lefevre, 1986). An emerging trend of interest concerns the *design* of assessment and forms of feedback that are alternative to grades or scores. There was consensus on the fact that assessment is a crucial part of the *teaching and learning process* and not a mere final "appendix". Concerning directions for further research, participants recognized the need for more research on classroom assessment practices and how to implement research results in real school contexts and in regular teaching practices. This brings to the fore the importance of researching assessment in collaboration with teachers. Connections between the contributions highlighted also the potentiality of comparative studies across countries. Finally, some emerging themes were recognized as crucial and worthy of further developments: the affective dimension of assessment and the role of inclusion, diversity and social justice in assessment.

References

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–28). Lawrence Erlbaum Associates, Inc.

Niss, M., & Højgaard, T. (2019). Mathematical competencies revisited. *Educational Studies in Mathematics, 102*(1), 9–28.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Erlbaum.