



HAL
open science

Proceedings of FAME 1 – Feedback & Assessment in Mathematics Education (ETC 14), 5-7 June 2024, Utrecht (The Netherlands).

Paola Iannone, Filip Moons, Christina Drüke-Noe, Eirini Geraniou, Francesca Morselli, Katrin Klingbeil, Michiel Veldhuis, Shai Olsher, Hankeln Corinna, Gonscherowski Peter, et al.

► To cite this version:

Paola Iannone, Filip Moons, Christina Drüke-Noe, Eirini Geraniou, Francesca Morselli, et al.. Proceedings of FAME 1 – Feedback & Assessment in Mathematics Education (ETC 14), 5-7 June 2024, Utrecht (The Netherlands).. FAME - Feedback & Assessment in Mathematics Education, Utrecht University & ERME, Jun 2024, Utrecht, Netherlands. 10.5281/zenodo.14231455 . hal-04807902

HAL Id: hal-04807902

<https://hal.science/hal-04807902v1>

Submitted on 6 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Utrecht
University



FAME 1 Proceedings

Feedback & Assessment in Mathematics Education

5-7 June, 2024 - Utrecht, The Netherlands



Edited by:

Paola Iannone, Filip Moons, Christina Drücke-Noe, Eirini Geraniou,
Francesca Morselli, Katrin Klingbeil, Michiel Veldhuis and Shai Olsher



european society for research in mathematics education

14th ERME Topic Conference (ETC14)

The International Programme Committee (IPC)

Chair of the IPC:

Paola Iannone (United Kingdom)

Members:

Christina Drüke-Noe (Germany)

Eirini Geraniou (United Kingdom)

Filip Moons (Belgium)

Francesca Morselli (Italy)

Katrin Klingbeil (Germany)

Michiel Veldhuis (The Netherlands)

Shai Olsher (Israel)

The Local Organising Committee (LOC)

Filip Moons (Belgium)

Michiel Veldhuis (The Netherlands)

Conference webpage:

<https://www.uu.nl/fame>

Editors: Paola Iannone, Filip Moons, Christina Drüke-Noe, Eirini Geraniou, Francesca Morselli, Katrin Klingbeil, Michiel Veldhuis and Shai Olsher

Publisher: Utrecht University

Place: Utrecht, The Netherlands

Year: 2024

ISBN: 978-90-393-7788-8

DOI: 10.5281/zenodo.14231455

© Copyright left to the authors

Recommended citation for the proceedings

Iaonnone, P., Moons, F., Drüke-Noe, C., Geraniou, E., Morselli, F., Klingbeil, K., Veldhuis, M., Olsher, S. (Eds.). (2024). *Proceedings of FAME 1 – Feedback & Assessment in Mathematics Education (ETC 14), 5-7 June 2024, Utrecht (The Netherlands)*. Utrecht University and ERME. <https://doi.org/10.5281/zenodo.14231455>

Recommended citation for single entries in the proceedings

Authors (2024). Title of paper/poster. In P. Iaonnone, F. Moons, et al. (Eds.). *Proceedings of FAME 1 – Feedback & Assessment in Mathematics Education (ETC 14), 5-7 June 2024, Utrecht (The Netherlands)* (pp. xx-yy). Utrecht University and ERME. <https://doi.org/10.5281/zenodo.14231455>

Summative Assessment across three countries

Francesca Morselli¹, Paul Drijvers², Alice Lemmo³, Hans-Georg Weigand⁴

¹ DIMA University of Genova, Italy; morselli@dima.unige.it

²Utrecht University, the Netherlands; p.drijvers@uu.nl

³Università dell'Aquila, Italy; alice.lemmo@univaq.it

⁴University of Würzburg, Germany; weigand@mathematik.uni-wuerzburg.de

Introduction

This article builds upon the themes discussed during the panel held at the conference, which focused on summative assessment. During the panel, three researchers—Drijvers, Lemmo and Weigand—presented information on three countries (Italy, the Netherlands, and Germany, respectively) and offered insights, reflections, and proposals for future directions. In preparation for the panel, a set of questions was shared with conference participants weeks in advance, allowing them to upload their responses to a Padlet platform.

This first section provides a brief summary of the key themes that emerged from the padlet. Researchers from six countries posted their contributions: China, Greece, Mexico, Singapore, Spain, and Turkey. Their comments helped to sketch a first overview of summative assessment practices, their purposes, and emerging trends in each context.

The subsequent sections present the contributions of the three experts.

First theme: how summative assessment is carried out

In most countries, summative assessment predominantly relies on written tests conducted with traditional methods, such as paper and pencil. These assessments are usually administered at the end of the semester or academic year, with final scores calculated as the average of results from multiple tests or assignments throughout the year. We may note that this approach may limit the exploration of alternative assessment methods.

Second theme: the purpose of summative assessment

The primary objectives of summative assessment across the surveyed countries include: ability streaming and entrance to schools and universities.

Summative assessment for ability streaming is used in contexts like Singapore: this practice aims to group students based on their academic abilities. However, research highlights potential drawbacks, such as stigmatization and a narrowing of curricula for students in lower ability streams (Kramer-Dahl & Kwek, 2010).

Summative assessment is used for entrance to schools and universities in countries such as Turkey, where summative evaluation scores play a critical role in determining student progression to higher education levels. Secondary school scores are used to assign students to high schools, while university placements depend on standardized entrance exams.

Third theme: emerging trends and key issues

The survey highlighted the following emerging trends and key issues related to summative assessment.

The first trend concerns equity and access. Several countries face disparities in assessment outcomes due to socioeconomic and geographical inequalities. For instance, students from rural or underprivileged urban areas in China often encounter additional challenges in accessing quality education. In Greece, efforts to create an inclusive public education system have led to a focus on evaluating students with learning difficulties under the general education framework, emphasizing the social dimensions of assessment.

The second trend is related to assessment stress and mental health. High-stakes assessments contribute to student stress and excessive test preparation. In response, countries like China are adopting new curriculum standards that emphasize holistic assessment and performance-based evaluations. Similarly, Singapore is exploring strategies to balance traditional examinations with broader educational goals, gradually reducing their weight while maintaining accountability.

The third trend concerns the use of Artificial Intelligence (AI). Spain has highlighted the growing influence of AI on educational assessments, particularly regarding the quality and integrity of student-submitted tasks. This technological shift raises questions about maintaining fairness and adapting assessment frameworks to the digital age.

The last trend refers to teacher professional development. Effective implementation of summative assessment reforms requires well-trained educators. However, countries such as Mexico report that teachers often lack the necessary training to adopt new methods, leading to the continued use of traditional approaches. Greece has also emphasized the need for professional development programs to support innovative assessment strategies.

Summative Assessment in the Netherlands

Paul Drijvers

The present state

Summative assessment at the national level in secondary education in the Netherlands has the form of final national examinations at the age of 16 years (vocational education), 17 years (pre-higher education), or 18 years old (pre-university education). For all school types, students' final grades are the average of their grade on the final national examination and the grade for school-based examinations, the latter having different possible formats depending on the schools' policy.

The national examinations for some tracks in vocational education are fully digital (assessment *through* technology, Olsher et al., 2023), but are graded manually by the teachers. For other school types, the final national examination is a paper-and-pen test, in which students can use a calculator or, in case of pre-higher and pre-university education, a non-CAS graphing calculator (assessment *with* technology).

Some reflections

The aim of the 2016 curriculum reform for pre-higher and pre-university education to give higher-order mathematical skills, called mathematical thinking activities, a more prominent place in both teaching and assessment has proven to be hard to implement in the final examinations (Drijvers et

al., 2019). How to assess such higher-order skills in the traditional assessment format, in combination with their positions in formative assessment, remains a challenge; a challenge that is not limited to the Netherlands but is widely recognized in our community (Olsher et al., 2023).

Further Developments

At present, a curriculum reform process for all educational levels is taking place in the Netherlands. Of course, the role of digital technology in summative assessment is part of the debate. The common opinion seems to be that graphing calculators are no longer up-to-date technology. One option is to switch to examinations through technology, or to examinations with digital tooling on a tablet or laptop in a so-called bootable client lockdown environment. Content wise, the question is whether or not to include CAS facilities, and if yes, to what school types. Once these questions will be answered and the new curricula will be implemented, the next question of how to deal with artificial intelligence as mathematical tools will be begging to be addressed as well.

Summative Assessment in Italy

Alice Lemmo

The present state

The aspects that principally characterize the Italian assessment policy concern teacher responsibility. The Italian national guidelines state that the responsibility for both periodic and final assessment lies with teachers (MIUR, 2012). In addition, the certification of competences at the end of middle (grade 9) and secondary school (grade 13) is also provided by the teachers.

The Italian standardised assessment (INVALSI) involves all grade 2, 5, 8, 10 and 13 Italian students. INVALSI test is in no way part of school assessment, but it is compulsory for participating in the final examination. It has no formal impact on the teaching-learning and assessing process, each teacher decides independently how to analyse, interpret and use the data collected from his/her students in teaching practice. INVALSI test from grade 8 to grade 13 is computer based, in the next two years this will also be the case for grade 2 and 5 (primary school). Italian standardised assessment therefore involves the use of technology (calculators, digital tools) and is administered through technology (computers/tablet), paper and pen are available if the student needs them.

At the end of middle and high school, all students take a final examination that consists in both written and oral assessment for all disciplines, including mathematics. For grade 8 students both written and oral test are designed and administered by teachers. The final mark consists in the performance in the exams, assessed by teachers. Differently, the final mark at the end of grade 13 is composed by both final exam performance and the assessments of the last three years of school career. In particular, 40% concerns the school career, the other 60% is divided into 1/3 for oral discussion and 2/3 for written examination, in turn divided into two tests. The two written exams are statewide, but teachers administer and evaluate them. Only the second test concerns mathematics, but it is administered only for the scientific high schools' students.

Some reflections

The main aspect of the Italian policy concerns the responsibility and autonomy of teachers in the assessment process at each school level. This could be interpreted because teachers' support guides

and assists students in their school career in the same manner as the assessment process should do in a formative perspective. Unfortunately, this may reveal as a point of deep weakness: such autonomy leads to several obstacles and critical issues, if it is not supported and trained.

The use of tests and written exams at school is becoming more and more common (Shepard, 2000); in Italy, written exams and tests seem to represent the main tool for class assessment in mathematics (Amado & Morselli, 2023). This tendency could be perceived as a remedy for inconsistencies in the evaluation or grading practices proposed to teachers and/or as a quest for evaluative objectivity (Shepard, 2000). My personal experience as teacher trainer revealed that Italian teachers seem confused formative and summative assessment with subjective and objective practices. This misconception endorses the assessment of procedural at the expense of conceptual understanding (Hiebert & Lefevre, 1986). This has a strong impact on teaching and learning whereby teaching to the test and consequently learning to the test are increasing. This is confirmed by grade 13 INVALSI test results. Students' data is presented divided in five levels: Level 1 represents students that only deal with elementary knowledge and procedural items, while Level 5 denotes students who tackle problem solving and argumentation items. About 50% of grade 13 students is in Level 1 and 2, only 15% of them belong to Level 5 (INVALSI, 2024). This is also confirmed by OECD PISA 2022 results: more than 50% of Italian grade 10 students belong to level 3 and below (OECD, 2023).

Further Developments

Several Italian researchers are engaged in research on formative assessment (e.g. Cusi, Morselli & Sabena, 2017); others on standardised assessment (e.g. Ferretti, Giberti & Lemmo, 2018). Little is yet done on summative assessment, in particular with tests or written exams (Lemmo, 2023). Concerning this last issue, Italian research is exploiting teaching-learning to the test, proposing tests designed upon processes and assessment grids that aim at competencies (Niss & Højgaard, 2019).

This perspective does not impose to drastically change tasks' design but to modify the way of assessing them. In a procedural approach, teachers could choose that the correct answer in a task counts a certain number of points, and he/she could subtract some points for each error or missing information. The feedback students and teachers receive focus on errors and/or missing answers. In a competency approach, each single task of the test contributes for identifying the level for each competency. For example, even with a wrong answer in some tasks, students could demonstrate ability in using algebraic calculations or properties; in other cases, they could show high communication competency and so on. In this latter case, the focus of the assessment is not on error or omission but on the level of the competencies the student employs.

In conclusion, there is a great gap between the development of research in mathematics education and assessment (Niss, 1992). Research needs to overcome this gap remembering that tests and written exams has an impact on the teaching-learning process and regulates it. In the Thirteenth Congress of the European Society for Research in Mathematics Education, into the Assessment in mathematics education working group, "there was consensus on the fact that assessment is a crucial part of the teaching and learning process and not a mere final "appendix" " (Morselli et al., 2023, p. 5).

Summative Assessment in Germany

Hans-Georg Weigand

The present state

Concerning the education system, there is no uniform system in Germany. There are 16 states, and each state is responsible for the education system on its own. The consequence is that there is a great variety of school systems and also of final examinations. Although there exist National Standards since 2004, and there is a national “task pool” for exams without technology (Fig. 1), the states can take different tasks from this pool for the first part of the final examination and the second part of the examination is completely different in different states. Concerning digital technologies there are states, which do not allow any digital technology in examinations (except arithmetic calculators), in some other states schools can choose whether they want to do the exam with advanced (usually scientific CAS-calculators) or without technology, and states where CAS-calculators are mandatory for all students.

The federal government tries to control the compliance of the national standards. There are different nationwide annual tests. VERA¹ is a comparative test in grade 3 and 9 (Fig 2 and 3). “IQB-Trends in student achievement”² is a test every 5 years in grade 4 and every 3 years in grade 9. It has been a comparative test between the states from 2008-2012, however since 2015 it should “only” give feedback to students, teachers and the policy makers.

Fig. 1 A task from the nationwide “task pool” – final exam

Function f .

Evaluate the following statement:

For each value of x with $0 \leq x \leq 2$, the slope of the graph of f is less than 3.

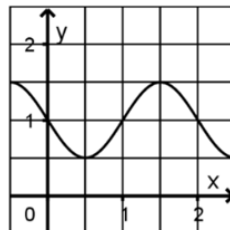


Fig. 2 VERA - grade 3: How many small boxes does the rectangle have?

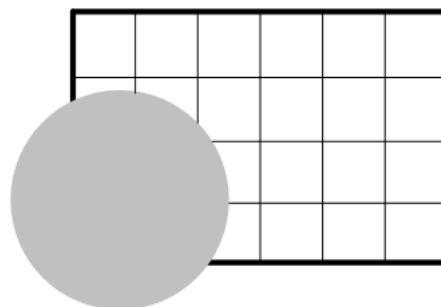


Fig. 3 VERA 8 – grade 9

¹ <https://www.iqb.hu-berlin.de/vera>

² <https://www.iqb.hu-berlin.de/bt>

Give reasons why there is no triangle with the sizes

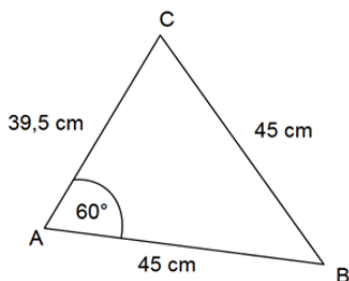


Fig. 4 Bavarian statewide test - grade 8

Simplify as much as possible!

a) $3x - 7y + 2y =$

b) $0,1b^3 \cdot 20b^2 =$

c) $\left(\frac{1}{2}x^2\right)^3 =$

German's reaction to the (bad) results of TIMMS, PISA, VERA, ...

There are – again – quite different reactions to the (bad) results of the international and national tests. E.g., Hamburg or Berlin assigned a scientific working group (2018, 2020). There have been many suggestions for changes concerning the goals, contents and classroom work and the states working on the implementation of some of these suggestions. Bavaria introduced statewide annual – quite classical – tests for grade 8 and 10 at the beginning of each school year (Fig. 4). Last year the nationwide professional development programme “QuaMath” for the next 10 years and for teacher of 10.000 schools was established and at the moment there is just starting a nationwide programme for low achieving students. Sometimes, reactions from the policy makers are done without any empirical basis. In Baden-Württemberg, since 2001 graphing calculators were allowed in examinations, however in 2014, there was a ministry ban and since then no advanced technology is allowed in final examinations. This is an example how arbitrary decisions concerning final examinations are taken.

Further Developments

There are some very important questions and emerging areas for further development concerning summative assessment that are discussed intensively (of course, not only) in Germany. Summative assessment ...

- ... has a crucial influence on classroom teaching (e.g., how digital media are used).

- ... cannot be a 1:1 copy of classroom work. There are some goals which are not – and cannot be – addressed in summative assessment, e.g., modeling or inquiry-based learning.
- ... must keep the balance between procedural and conceptual understanding (CU).
- ... but must be seen in relation to the development of adequate tasks for CU in the classroom.
- ... requires follow-up research on the reasons for good/poor performance.
- ... needs strategies to react – in the short and long term – to assessment results.
- ... needs to think about alternatives, e.g., oral exams, project work or portfolios.
- ... has to be open for up-coming developments like virtual reality, augmented reality or artificial intelligence.

References

- Amado, N., & Morselli, F. (2023) Teachers' beliefs about assessment: a study in Italy and Portugal. In In Drijvers, P., Csapodi, C., Palmér, H., Gosztonyi, K., & Kónya, E. (Eds.). (2023). *Proceedings of the Thirteenth Congress of the European Society for Research in Mathematics Education (CERME13)*. Alfréd Rényi Institute of Mathematics and ERME. <https://hal.science/hal-04413032v1>
- Cusi, A., Morselli, F. & Sabena, C. Promoting formative assessment in a connected classroom environment: design and implementation of digital resources. *ZDM Mathematics Education* 49, 755–767 (2017). <https://doi.org/10.1007/s11858-017-0878-0>
- Drijvers, P., Kodde-Buitenhuis, H., & Doorman, M. (2019). Assessing mathematical thinking as part of curriculum reform in the Netherlands. *Educational Studies in Mathematics*, 102(3), 435–456. <https://doi.org/10.1007/s10649-019-09905-7>
- Ferretti, F., Giberti, C., & Lemmo, A. (2018). The Didactic Contract to interpret some statistical evidence in mathematics standardized assessment tests. *EURASIA Journal of Mathematics, Science and Technology Education*, 14(7), 2895-2906. <https://doi.org/10.29333/ejmste/90988>
- 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). Routledge <https://doi.org/10.4324/9780203063538>
- Kramer-Dahl, A., & Kwek, D. (2010). 'Reading' the home and reading in school: Framing deficit constructions as learning difficulties in Singapore English classrooms. In C. Wyatt-Smith, J. Elkins, & S. Gunn (Eds.), *Multiple perspectives on difficulties in learning literacy and numeracy* (pp. 159–178). Dordrecht: Springer.
- INVALSI (2024). *Rapporto Invalsi 2024*. Retrieved from https://invalsi-areaprove.cineca.it/docs/2024/Rilevazioni_Nazionali/Rapporto/Rapporto%20Prove%20INVALSI%202024.pdf
- Lemmo, A. (2023). Mathematical written tests as formative assessment practice. In In Drijvers, P., Csapodi, C., Palmér, H., Gosztonyi, K., & Kónya, E. (Eds.). (2023). *Proceedings of the Thirteenth*

Congress of the European Society for Research in Mathematics Education (CERME13). Alfréd Rényi Institute of Mathematics and ERME. <https://hal.science/hal-04413550>

MIUR (2012). *National guidelines for the curriculum for pre-school and first cycle of education*. [National guidelines for the curriculum for grades K-8]. *Annali della Pubblica Istruzione*, Numero Speciale. Le Monnier.

Morselli, F., Drüke-Noe, C., Giberti, C., Kaplan-Can, G. & Rämö, J. (2023). An introduction to TWG21: Assessment in mathematics education. *Proceedings of the Thirteenth Congress of the European Society for Research in Mathematics Education (CERME13)*. Alfréd Rényi Institute of Mathematics and ERME. <https://hal.science/hal-04412990v1>

Niss, M. (Ed.). (1992). *Investigations into assessment in mathematics education: An ICMI study* (Vol. 2). Springer Dordrecht. <https://doi.org/10.1007/978-94-017-1974-2>

Niss, M., & Højgaard, T. (2019). Mathematical competencies revisited. *Educational Studies in Mathematics*, 102, 9-28. <https://doi.org/10.1007/s10649-019-09903-9>

OECD (2023), PISA 2022 Results (Volume I): *The State of Learning and Equity in Education*, PISA, OECD Publishing, Paris, <https://doi.org/10.1787/53f23881-en>.

Olsher, S., Chazan, D., Drijvers, P., Sangwin, C., Yerushalmy, M. (2023). Digital assessment and the “machine”. In B. Pepin, G. Gueudet, & J. Choppin (Eds.), *Handbook of Digital Resources in Mathematics Education*. Springer. https://doi.org/10.1007/978-3-030-95060-6_44-1

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14. <https://doi.org/10.3102/0013189X029007004>

FAME 1 Proceedings
Feedback & Assessment in Mathematics Education
5-7 June, 2024 - Utrecht, The Netherlands

14th ERME Topic Conference (ETC14)

ISBN: 978-90-393-7788-8
DOI: 10.5281/zenodo.14231455



**Utrecht
University**