DG 12: Assessment and testing shaping education, for better and for worse

Team Chairs: ***Glenda Lappan***, Michigan State University, East Lansing, USA
Dylan Wiliam, Learning and Teaching Research Center, ETS, Princeton, USA
Team Members: ***Sean Close***, St. Patrick's College, Dublin, Ireland
Gabriele Kaiser, University of Hamburg, Germany
Vijay Reddy, Human Sciences Research Council, South Africa

Aims and focus

This Discussion Group focused on the impact that educational assessments have on the teaching of mathematics, and was organized around three main themes: what mathematics should be assessed; the alignment of standards, curriculum and assessment; and how mathematics should be assessed.

What mathematics should be assessed?

One of the key areas of agreement in the three sub-groups (organized around the three themes) was that the intended goals and curriculum (the mathematics expectations or standards) should determine the way assessment in mathematics is done, rather than the other way round. Therefore, the connection between these should be clear to all – including students, teachers, parents, and other interested agencies.

The focus of assessment should be expanded to include both content and process aspects of mathematics (see paper by vom Hofe et al on the DG 12 web-site). Assessment of content alone without regard to mathematical processes (e.g. communicating, modeling, reasoning, showing rigor in approach, and connecting) is not sufficient. On the other hand, processes do not exist independently of curricular content – for example, if students are to communicate, they have to communicate about something. However, assessing mathematical process is not easy since exemplary problems that have the potential to promote the development of mathematical processes in students can, in the hands of some teachers, be reduced to a set of routinized skills. For example, in England, where the assessment of mathematical investigations was introduced into the national school leaving examination in the 1980s, in many classrooms this was taught as an additional piece of content, with students given a set of procedures to follow. Similarly in Ontario, Canada (see paper by Suurtamm and Lawson on the DG 12 web-site), a new curriculum emphasized instruction using challenging problems, the student construction of multiple solution methods, and mathematical communication and defence of ideas. However, the curriculum was operationalized through a detailed list of content standards that teachers were expected to cover, which militated against the rich curriculum envisaged.

Results from international attempts to assess mathematical processes reveal that worldwide the emphasis on processes is not as well developed as is needed for students to become users of mathematics in real contexts. Progress in this area can begin with helping teachers design and carry out assessments of mathematical process in their classrooms. Such work can also have the side effect of helping teachers develop their teaching practices and learn to use assessment tasks that promote process aspects of mathematics.

Of course, these process elements of mathematics should permeate all the content aspects of mathematics, but there is the separate, essentially strategic, issue, of whether these process aspects should be identified separately. In the development of the NCTM 1989 *Curriculum and Evaluation Standards for School Mathematics K–12*, the writing groups did articulate standards that were more like traditional content standards but added a set of standards – communicating, problem solving, reasoning and connecting – that were meant to convey the importance of these bigger more powerful competencies. Similarly, in England, the National Curriculum for mathematics specifies traditional content areas (number, shape and space, probability and statistics) as well as separate process aspects, and process aspects are also clearly identified within the PISA assessment framework. While separating content and process in this way does create difficulties, it also highlights the importance of process aspects of mathematics. Ideally, the process aspects should permeate all the content aspects of mathematics curricula, but initially at least it may be necessary to signal the importance of process aspects by making them separate.

Another issue that emerged in many discussions was that of curricular progression. In many countries, the connections between the curricula for different years or grades are very weak, so that the fundamental question of progression ("When someone gets better at mathematics, what is it that gets better?") is rarely addressed. This is exacerbated in countries, such as the USA, where teachers tend to teach only one grade or year, so that while a teacher may, for example, understand the fourth-grade mathematics curriculum well, they have only a hazy idea of how this builds on the third-grade curriculum, and how it leads on to the fifth-grade curriculum.

Alignment of curricula, teaching and assessment

In many countries, (e.g. France, Japan, Sweden) the mathematics curriculum is the responsibility of central government, while in others (Australia, Germany, United States) this is devolved to regional bodies (states or regions). In some countries, the mathematics curriculum is the responsibility of the district, the school, or even the individual teacher. However, even when the curriculum is clearly specified, in many countries the assessments are not well aligned with the curriculum. Two problems are particularly common. The first is *construct under-representation*. This occurs when the assessment systematically under-represents the construct of mathematics as defined in curriculum standards (for example when an assessment does not assess process aspects of mathematics even though this is an explicit component of the curriculum). The consequence of this is that where schools are under pressure to improve their students' results, those aspects of mathematics that are not assessed are neglected. As a result the curriculum does come to be aligned with the assessment, but only by changing the curriculum (instead of making the important mathematics measurable, we end up making the measurable mathematics important).

The second problem is that of *construct-irrelevant variance*. This occurs when the assessment requires students to have skills or knowledge beyond those that it is intended to assess. For example a mathematics test might be couched in complex language so that to succeed, the student would need to be good at mathematics and reading. When a student gets a low score, we do not know whether this is because they cannot do the mathematics, or whether they could not understand what they were being asked to do (in the latter situation, part of the variance in scores is attributable to factors not relevant to the construct being assessed).

From the discussions in DG 12, it appeared that lack of alignment between curriculum and assessment appeared to be a widespread, if not universal problem, although in some countries there were signs of progress. In France, new methods are being developed for assessing 'process' aspects of mathematics in the Baccalauréat such as the quality of exposition, accuracy of justifications and coherence in the reasoning process (see paper by Feurly-Reynaud on the DG 12 web-site). In Sweden there have been considerable changes in the mathematics tests given to 18-year-olds over the last 30 years (see paper by Jakobsson-Åhl on the DG 12 web-site). In 1973, the national test was abstract, placing great emphasis on decontextualised, axiomatic thinking (the result of the introduction of the "new math" in Sweden at that time). The 2002 test reflects more of an emphasis on student thinking, allowing more time and including three lengthy "story" problems. The exams were also scored differently, with the 2002 requiring complete solutions, rather than just answers. In the Netherlands, the development of investigative skills was introduced into the mathematics curriculum for lower secondary schools in 1993, but the national tests were traditional pencil-and-paper tests, which did not assess adequately the investigative aspects of the curriculum (see paper by Vos on the DG 12 web-site). To remedy this, the National Institute for Educational Measurement developed a series of practical tests so that even those teachers who saw their role as 'teaching to the test' had an incentive to incorporate investigative work into their teaching.

However, even when the assessments are well-aligned with the curriculum, both can be seriously out of alignment with the teaching in classrooms. In Portugal (see paper by Carvalho e Silva on the DG 12 web-site), two committees were established in the mid 1990s, one to work on assessment and one to work on curriculum. A subsequent evaluation of their work established that they did manage to align the assessment with the curriculum to a good degree, but that test items specifically aligned to the standards (e.g. using graphing calculators to find the intersection of 2 curves) were among those with the worst student performance. In other words, the assessment was well-aligned with the 'intended' curriculum' but not with what the students actually learn (the 'achieved' curriculum), showing that large-scale assessment does not necessarily drive practice. A similar outcome was found in South Africa where the new 'Common Task Assessments', which focused on students' ability to use their mathematics to solve 'real' problems, were perceived by teachers to be neglecting the more 'academic' mathematics the students had studied (see paper by Naidoo and Parker on the DG 12 web-site). Such examples show that aligning curriculum standards with assessment is only the first step. We need also to find better ways of communicating about these curricular standards to practitioners, in particular for the process standards. We also need better examples of problems and rubrics for assessing student performance.

How should we assess mathematics?

It was agreed that the predominant methods of assessing mathematics, such as timed written tests and examinations, do not assess adequately many of the aspects of mathematics curricula around the world. This is not to say that such tests and examinations are useless, but, especially where curricula give substantial emphasis to mathematical processes, other forms of assessment will also be necessary.

Since many of these process aspects of mathematics can be assessed only through extended pieces of work, two things follow. The first is that in order to justify the time taken for such extended pieces of work, the assessment tasks must also be tasks during which students learn (assessment *as* learning). The second is that if the costs of assessments are to be kept down, teachers must be involved in summative assessment, so that the assessments that teachers make as part of their normal work can contribute to the overall judgments made about the capability of students.

To achieve this, teachers will need examples of exemplary tasks, rubrics for scoring such tasks, and professional development focused explicitly on teaching practices that support the use of such tasks and the development of the appropriate mathematical competencies in students. The good news is that there is a range of models in use around the world that can provide some ideas about how to broaden the basis of assessment. Denmark, and many former Soviet bloc countries have a strong tradition of oral examination in mathematics. In Sweden, there are national tests in mathematics at age 18, but these are used only to guide the teacher's judgment of the student's grade, which is based on all the work done in the final years of upper secondary school (students who get low grades can also get into university on the basis of a special aptitude test). In Queensland, Australia, again, university entrance is based on teacher judgments but there is a 'core skills test' that is used to calibrate performance across different subjects. In the USA, practice is very varied, but assessment of mathematics appears to be conducted primarily for the purpose of holding schools to account for the performance of their students, so that the assessments tend to be low-stakes for the students but high-stakes for the schools. In order to keep costs down, most states use standardized multiple-choice tests, and while it is possible to measure some higher-order skills in this way, it seems fair to say that the tests in widespread use do not. Furthermore, many tests are kept secret so that teachers have only a hazy idea of the content of the tests their students will take.

Summary of DG 12 discussions

As might be expected, the group came to no clear consensus about the best ways to assess mathematics. Indeed, a clear finding of the group was that the assessment systems in place in each of the countries are inextricably linked to the local contexts in which the systems operate. Solutions that work in one setting are unlikely to work in another without some adaptation. However, there appears to be general agreement about some broad principles that should govern all assessment systems.

- Education goals and curriculum expectations should determine the way assessments are done.
- Teachers are a critical part of the assessment process.
- The process should be as transparent as possible, meaning that students, parents, teachers, administrators, and school boards should all be aware of what the tests will comprise and be supportive of the goals.
- There is a need for greater flexibility in the timing of exams.
- In order to test what we care about, including thinking and reasoning, we need more innovative assessment tools and accompanying text materials.
- We need to ask what the functions of assessment are. – Why assess?
- We need to ask what the domain of assessment is. – What to assess?

- We need to ask who has authority for assessment. – Who is involved and why?
- What is the target of assessment? – Who is assessed and why?
- What is the means of assessment? – How to assess?
- What are the constraints and the affordances of the testing process. – What does a test reveal to students and teachers that helps improve mathematics education?

This report was written by Glenda Lappan (glappan@math.msu.edu) and Dylan Wiliam (dylanwiliam@mac.com) who are happy to be contacted for further information on the work of this DG.