

DG 13: Evaluation of teachers, curricula and systems

Team Chairs: *Claude Gaulin*, Laval University, Québec, Canada
Max Stephens, University of Melbourne, Australia
Team Members: *Hugh Burkhardt*, University of Nottingham, United Kingdom
Einar Jahr, Hedmark College, Hamar, Norway



I C M E
1 0
2 0 0 4

DG

Discussion
Group 13

Aims and focus

All participants agreed that the theme raised key issues for policy and practice, locally, nationally and internationally, and that simple recipes for evaluation are to be avoided. Three sub-groups corresponding to the key words “teachers”, “curricula”, and “systems” then met and reported back on key questions and findings. It was agreed that focusing on the quality of students’ learning is a necessary link across all three areas of evaluation.

Evaluation of teachers

Several countries are using professional teaching standards as a tool for individual and systemic evaluation of teachers of mathematics. Public, credible standards – and the assessment of individuals against these – are fundamental to the high esteem in which other professions are held. How these other professions practise self-regulation enhances their standing, and adapting this model has the potential to produce the same positive outcomes for teachers of mathematics (Bishop, Clarke, & Bennett, 2000; Ingvarson, 1998). Some significant claims have been made about the power of standards and their assessment:

The National Board for Professional Teaching Standards is ... dedicated to student learning and to upholding high standards for professional performance. We have raised the standards for teachers, strengthened their educational preparation through standards, and created performance-based assessments that demonstrate accomplished application of the standards.
(NBPTS, www.nbpts.org/about/index.cfm)

However, teaching is a complex performance and there are many challenges in assessing accomplishment (Delandshire & Petrovsky, 1999). A number of tensions arise for teachers as they develop and present evidence of their accomplishments:

- *Comparability versus creativity* – Can I exercise creativity in presenting evidence of my accomplishments to assessors who need to show consistency in their judgements?
- *Meaningful versus manageable* – How can I give evidence of my accomplishment in ways that do justice to my achievements and yet avoid excessive amounts of time in assembling and presenting evidence? How much evidence do assessors really need? What do I tell them so that they fully understand the nature of my achievements?
- *Accomplishment versus ongoing critical inquiry* – To what extent can I use this process as a means of engaging in genuinely critical inquiry into my own teaching, rather than simply demonstrating my accomplishments?
- *Personal goals versus school policy* – Is it possible to demonstrate personally meaningful accomplishments beyond the professional development goals required by my school or those required by government policy?



I C M E
1 0
2 0 0 4

The following questions guided group discussion on the evaluation of teachers.

- What are the “costs” and benefits of evaluating teachers? If all teachers are to be evaluated, does this encourage better teaching or distort priorities?
- What methods are available – what are their strengths and weaknesses?
- How might teacher self-evaluation be developed?

There was considerable discussion trying to answer the question “What is it that it is most important to evaluate?” The resulting list of four points has many similar themes to “standards documents” developed in, for example, Australia and the USA:

1. *Subject matter knowledge for teaching mathematics* – including both breadth and depth of knowledge, clear understanding of the connections between concepts, an understanding of potential difficulties students may have, use of appropriate representations and knowledge of typical misconceptions students may have.
2. *Knowledge of student learning* – including the need to provide opportunities for all students to learn, knowing how children learn, and what they have learnt as well.
3. *Professional growth* – including a commitment to their own learning and being open to reflection – to keep reassessing their own practice.
4. *Teaching practice* – successful planning and implementation of appropriate tasks, developing a culture and environment where learning happens.

In considering these four points, the group strongly agreed that any evaluation process should include evidence of *actual* student learning – have students learnt some important mathematics and does a teacher facilitate independent thinking of students?

As well as the complexities of evaluating teaching, there are the multiple agendas of those involved. Resolving or just managing the professional, the political and the systemic contexts is difficult. But a key point was the need to resist simplifying both a framework for describing teaching and also the process of evaluating it. Teaching is a complex activity and any evaluative process developed must both acknowledge and work within that complexity. While much of the initiative for teacher evaluation comes from systems, the voice of practitioners – teachers of mathematics – must lead the way in which standards for teaching are conceptualized, developed and implemented.

Evaluation of curricula

It is insufficient to evaluate curricula by inviting comments from, for example, an expert panel. The discussion group argued for a more rigorous process of gathering and using evidence of the implementation of a curriculum; its impact on the work of teachers and students, including the extent to which students and schools can demonstrate achievement of standards embodied in an intended curriculum.

Evaluation is needed to provide feedback on the implementation of a curriculum, including the impact of curriculum reform, in a range of realistic contexts affecting school personnel and resources. This information is necessary to revise (or reform) a curriculum in the *next* stage by evaluating *current* school practice. Such feedback is also necessary to provide rich and reliable information to guide policy and decisions regarding textbooks; and to provide reliable insights into the quality of teaching and learning.

DG

Discussion
Group 13



I C M E
1 0
2 0 0 4

DG

Discussion
Group 13

The group's view of curriculum includes not only statements of content and their relation to different year levels, but *also* recommendations on how the curriculum is to be taught; for example, through sample programs or advice to teachers. It was also argued that any evaluation of the curriculum must include the procedures that are used to assess students' performance either externally, or by internal classroom assessment.

Any rigorous evaluation of a curriculum must therefore attend to the full set of documents and procedures that are intended to explain and give effect to a curriculum. There will be room for interpretation, especially in the case of curricula that operate across a large school district or a whole state, and even more so for a national curriculum. While different implementations will be encouraged, some *givens* must be spelled out clearly and unambiguously in terms of what students are expected to know and to do.

The group argued that several other *givens* must be present. (1) A curriculum must contain a clear description of content to be covered, how that content is to be distributed across the various year levels, and how elements of content are linked together year by year. (2) A published curriculum has to spell out a comprehensive set of *standards* which relate to students' attitudes towards mathematics, the forms of mathematical thinking that are to be encouraged, the skills students are to have, how students are to represent mathematics, and what mathematical understanding they are to achieve. (3) It must be clear to those implementing a curriculum how they can be sure that all the intended standards have been covered. (4) Finally, any published curriculum must provide clear guidance for the development, appraisal and adoption of textbooks and other resources.

Evaluating an implemented curriculum must therefore have a clear focus on the *quality of learning* by students. Consequently, any well-designed curriculum must contain a range of teaching and assessment tasks exemplifying a range of formats and task demands. Actual student responses should be used where possible to illustrate the range and *quality* of successful performances, common misconceptions and errors. These are necessary to inform teaching, to guide the construction of assessment tasks, and so ensure that teaching and assessment address all aspects of the intended curriculum. They also show how implementation of a curriculum will be evaluated.

Evaluating the implementation of a curriculum can take place at the level of an individual school, at local or district level, or at state or national level. One example of a large-scale evaluation of the implementation of the curriculum is the extensive surveying of students' mathematical performance carried out periodically by the Ministry of Education in Japan. In evaluating implementation of a curriculum, it is possible to use innovative and experimental tasks that may not be appropriate for measuring school accountability or testing student achievement. Thus, the Japanese surveys include questions to elicit students' attitudes to mathematics generally and to particular topics. Two important results flow from evaluating a curriculum in this way. In the first place, tasks can be used to challenge current assessment and to change teaching in areas that may not be aligned with the intended curriculum. Second, performance standards can be established for students at specific year levels, and indicate how performance can be expected to change over time.

Evaluation of systems

Evaluation of systems is judging the value and worth of systems, their actions, and their conditions to produce quality learning in mathematics. The group agreed that a system



I C M E
1 0
2 0 0 4

DG

Discussion
Group 13

includes all who have responsibility to provide education, students, materials, conditions, forces, and other influences. A system also includes interconnections and interactions between these elements for the general purpose of serving the mathematics education needs of students, consumers of education, the community, and society at large.

A system under consideration for an evaluation generally is a school or a school district, but it could be a classroom, a state, or even a nation. It was found helpful to identify four components in thinking about designing an evaluation of a system. These are Student performance, Program, Policy and Contextual factors. *Student performance* constitutes the main aim of any evaluation. The *Program* consists of teachers, their practice, curriculum, materials, professional development, and other factors intended to lead towards desired student performances. *Policy* comprises the decisions and directives made to guide the system. Policy can be made at all levels within the system. Finally, there are *contextual factors* that may not be under the control of those in the system being evaluated, but are still influential on what the system does and what it accomplishes. The needs of the work force, higher education entry requirements, and parent expectations are such factors. All four components need to be built into an evaluation plan of a system.

The discussion noted that three types of variables – enabling variables, target variables, and explanatory variables – should be part of any model for a system evaluation.

Target variables represent elements that are specifically identified to be changed by a system initiative. Improved student performance is generally the primary target. But other target variables can include curriculum, alignment, saturation, equity, and quality. These variables can be seen as means or pathways to achieving higher student achievement. Having a more aligned system with higher quality curriculum and professional development is thought to lead to improved student performance.

Enabling variables represent those conditions that can either inhibit or facilitate the attainment of the target variables. These variables or conditions are necessary for advancing student performance within the system. Capacity and sustainability are two examples of enabling variables. For a system to attain the desired student performance and to support a new initiative requires the system to have some capacity including resources, social capacity, and other factors that can deliver a desired change in the system. Any initiative needs to be sustained adequately to have an impact. For example, professional development needs to reach all teachers in need to have a full impact on the system.

Explanatory variables help explain why or why not the target variables reach the desired goal. The group noted that when a new curriculum is adopted, some *trade offs* are inevitable. When new topics are added to the curriculum, other topics are eliminated. Identifying *trade offs* can explain why student performance in one area can increase while performance on another topic decreases. Cost and resource allocations can help to explain why desired goals are not being obtained or why an initiative has been successful.

At a minimal level, the evaluation of a system must include some measures of student performance. This measure should include both an indication of student growth and attainment over time. Ideally the measure of student growth will be an assessment of the same cohort of students at the beginning of the school year and at the end of the school year. The students who are assessed can be a sample of the population. In addi-

tion, there has to be some measure of practices within the system including variables from all three categories. It may not be necessary or possible to measure all three types, but the more variables that can be measured, the more precise the evaluation information will be.

Three issues related to evaluation of systems were identified. The first is to explicate *what the system is* in terms of the four components noted above, and defining expectations for how the system should function. A second issue is *time* for measuring the impact of some initiative. Trying to measure the impact prematurely can produce discouraging results and cause the discontinuing of resources for an initiative. Such poor results could be a result of underestimating the time when impact can first be expected. A third issue is the problem of *attribution*. Since systems are complex it is difficult to isolate specific factors and attribute change in the system to these factors. If the evaluation is to produce information on reasons why the system has changed, the major challenge for any evaluation is to identify the variables or initiatives that have contributed to the observed change.

References

- Bishop, A. J., Clarke, B., & Bennett, S. (2000, July). A research and development project on national professional standards for excellence in teaching mathematics. Paper presented to working group 7: The professional pre- and in-service education of mathematics teachers. Ninth International Congress on Mathematical Education, Tokyo, Japan.
- Delandshere, G., & Petrosky, A. (1998). Assessment of complex performances: Limitations of key measurement assumptions. *Educational Researcher*, *27* (2), 14-24.
- Ingvarson, L. (1998). Professional development as the pursuit of professional standards: The standards-based professional development system. *Teaching and Teacher Education*, *14* (1), 127-140.

This report was written by Max Stephens in cooperation with group members Barbara Clarke and Norman Webb who wrote draft reports of sections on “teachers” and “systems” respectively. Max Stephens can be contacted at m.stephens@unimelb.edu.au for further information on the work of this DG.



I C M E
1 0
2 0 0 4

DG

Discussion
Group 13