

Assessment for Improvement

Hugh Burkhardt*

MARS: Mathematics Assessment Resource Service

These notes are based on a presentation at the May 2004 meeting of Principal Investigators of Instructional Materials Development projects, funded by the National Science Foundation.

The aim is to offer a picture of:

- the various roles that assessment can play in forwarding improvement in teaching and learning in schools, complementing and supporting the roles of instructional materials;
- assessment tools that are available for mathematics, notably those that have been developed by MARS.

together with a brief discussion of the principles underlying the design of balanced assessment that is aligned with the NSF-funded curricula.

How can assessment help curriculum improvement?

Those active in reform are so used to assessment that it is a barrier to what they are trying to achieve that they ignore its power, when designed appropriately, to support progress. The potential of such an effective lever is worth exploring and exploiting. The following are some of the ways in which assessment can be used to advance reform, and the adoption of a good standards-based curriculum.

- **High-stakes assessment** that is aligned with the NCTM standards provides strong motivation for schools to adopt curricula that are similarly aligned. Traditional tests, like traditional curricula, cover a much narrower range of mathematical performance. This limited range excludes, for example, non-routine problem solving and other investigative activities. Broadening high-stakes assessment involves major challenges at policy level in school districts and states – but the payoff is comparably large (WYTIWYG). NCLB may actually help here – it requires state tests to be aligned with state standards which, having been written by curriculum people are often quite broad.
- **Curriculum evaluations** should always use instruments that cover the full range of the curriculum's goals. For NSF curricula this means broad and balanced assessment. Studies of this kind have shown massive gains on such tests along

* Michigan State University and the Shell Centre, School of Education, University of Nottingham, England NG8 1BB <Hugh.Burkhardt@nottingham.ac.uk>

with small gains on traditional tests. ("Decathlon" performance cannot be fairly judged on the 100 meters time alone. If it were, the effect on the training program is likely to be severe)

- **Classroom assessment** that involves students showing their working and explaining their reasoning provides valuable feedback to guide future learning and teaching. (Answers to multiple choice items reveal little)
- **Performance tasks illustrate curriculum goals** in a compact way that communicates vividly to teachers and students. (Curriculum materials are so substantial!) The scoring scheme for each task makes explicit what aspects of performance are valued, helping to define the subject more clearly. Samples of student work show the patterns of difficulty and success – and make unfamiliar challenges seem accessible. They are thus important elements in >
- **Professional development for and through assessment** can make a powerful contribution to any professional development program through task-focused discussion of various kinds. The feedback that assessment represents is a key element in helping teachers develop the strategies and skills of a student-focused teaching style.

MARS contributions

The MARS team has been funded by NSF since 1992 to develop and disseminate balanced assessment that is aligned with NCTM and international standards, in parallel with and in support of the curriculum development projects. Since then, working with school systems across the US and elsewhere, it has developed a range of products including:

- **Tests for Grades 3 through 10**, designed by MARS and published annually by CTB McGraw Hill under the title *Balanced Assessment in Mathematics*. Secure 'scaled' versions of the tests are also available for Grades 3-5, and will be soon for Grades 6-8. These are also used for program evaluation.
- **2 books of materials for each grade range** for use in classroom assessment and instruction, published under the same title by Dale Seymour/Pearson Learning;
- **Professional development support** materials, and workshops based on them, are published by MARS at Michigan State. These annual workshops, held each June, are an excellent way to learn about the power of good assessment in supporting reform.

[The materials have been developed to 'cascade' successfully, ie workshop participants who go on to use them to lead workshops in their own systems are seen to sustain quality comparable to the original. This rare outcome is based on good 'engineering' – outstanding design and systematic development of the materials through trials.]

- **Assessment consulting** to school systems engaged in reform. This work has revealed the broad range of challenges that change agents in school systems meet, leading to the current project to develop a *Toolkit for Change Agents*.
- **Other products** include, for example, a set of "*World Class Tests*" of *problem solving in mathematics, science and technology*, with associated teaching and learning modules, each focused on a specific problem solving strategy. These hold promise for the assessment of "science process" skills

More can be found on each of the above from www.msu.edu/MARS/, or www.mathshell.com, or by sending an inquiry to Shell.Centre@nottingham.ac.uk.

Some other fine materials are available. The New Standards Mathematics Reference Examinations provide a comprehensive standards-based assessment at Grades 4, 8 and 10. The International Baccalaureate provides challenging examinations for the end of high-school.

Design principles for assessment

The following brief notes set out some of the key issues in the design of assessment aligned with standards that address the broad range of aspects and types of performance needed to do mathematics or science well.

The Roles of Assessment

Assessment fulfils a number of important roles in the specifying, teaching and learning of mathematics K-12:

A: to 'measure' performance – ie

"to enable students to show what they know, understand and can do"

B: to exemplify the performance goals

also, with high-stakes, *inevitably*

C: to drive classroom learning activities

WYTIWYG (What You Test Is What You Get)

We are so used to the negative effects of standardized tests on learning and teaching in the classroom that the very idea seems absurd; yet the power of assessment can help (Clarke and Stephens on the Victoria, Australia reforms)

These roles carry responsibilities which designers should accept

Often they do not – arguing, for example, that the results of narrow tests "correlates well" with broader and more balanced (and more expensive) assessment. This argument addresses only does A, with dangerous side effects under B and C ("Mathematics is short multiple choice items and we should focus on them in our teaching") So **correlation is not enough.**

Key elements in reform

It is worth reviewing the main tools that are needed to successfully forward a standards-based improvement program. We need both tools and networks.

Key tools include exemplars, then substantial sets of:

- performance tasks that span the goals
- curriculum units
- professional development support units
- well-aligned tests for accountability

always with networking that builds

- public consensus building
- a good 'case'

Balanced Assessment

is assessment that accepts its responsibilities – A, B and C

Key design principles are:

- **curriculum balance** – tests should be designed such that teachers who "teach to the test" are led to provide a rich and balanced curriculum
- **learning value** – balanced assessment takes time, so doing each assessment tasks should be a good learning activity.

This is met most easily if the assessment tasks cover the performance goals of the curriculum in a balanced way. What is **not** balanced? For example: assessment that:

- only assesses number, no geometry
- only assesses concepts and/or skills
- has no non-imitative tasks (problem solving, applications)
- only has short 'items'
- is all multiple-choice

Cost

It is important to identify and feature the core question:

What is appropriate expenditure on assessment?

not, What do we now spend on it?

Assessment is a principal feedback channel in student learning

Feedback is crucial or any complex interactive system; it is common to spend

~10% turnover on 'instrumentation' and the feedback it produces

mostly formative, but including summative. Estimates suggest:

turnover ~ \$6,000 a student-year

total assessment: \$600 /yr

summative ~ \$60 /yr

Toolkit for Change Agents

"Change is made by people, but people are much more effective when they have well-engineered tools to use."

This is a current MARS team development. We aim to:

identify challenges that change agents face in their reform programs

collect tools that others have used successfully in overcoming similar challenges and, where necessary, develop them so that others can use them.

Assessment provides a part of the toolkit, offering some powerful tools to support reform. Curriculum materials are another section. Other tools* are less familiar.

We are working with change agents across the US. If you are interested, contact us at www.toolkitforchange.org

* For example, a *Math Wars Disarmament Tool*.