

Firm evidence shows that formative assessment is an essential component of classroom work and that its development can raise standards of achievement, Mr. Black and Mr. William point out. Indeed, they know of no other way of raising standards for which such a strong prima facie case can be made.

Raising the standards of learning that are achieved through schooling is an important national priority. In recent years, governments throughout the world have been more and more vigorous in making changes in pursuit of this aim. National, state, and district standards; target setting; enhanced programs for the external testing of students' performance; surveys such as NAEP (National Assessment of Educational Progress) and TIMSS (Third International Mathematics and Science Study); initiatives to improve school planning and management; and more frequent and thorough inspection are all means toward the same end. But the sum of all these reforms has not added up to an effective policy because something is missing.

Learning is driven by what teachers and pupils do in classrooms. Teachers have to manage complicated and demanding situations, channeling the personal, emotional, and social pressures of a group of 30 or more youngsters in order to help them learn immediately and become better learners in the future. Standards can be raised only if teachers can tackle this task more effectively. What is missing from the efforts alluded to above is any direct help with this task. This fact was recognized in the TIMSS video study: "A focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve."¹

In terms of systems engineering, present policies in the U.S. and in many other countries seem to treat the classroom as a black box. Certain inputs from the outside -- pupils, teachers, other resources, management rules and requirements, parental anxieties, standards, tests with high stakes, and so on -- are fed into the box. Some outputs are supposed to follow: pupils who are more knowledgeable and competent, better test results, teachers who are reasonably satisfied, and so on. But what is happening inside the box? How can anyone be sure that a particular set of new inputs will produce better outputs if we don't at least study what happens inside? And why is it that most of the reform initiatives mentioned in the first paragraph are not aimed at giving direct help and support to the work of teachers in classrooms?

..... **The Argument**

We start from the self-evident proposition that teaching and learning must be interactive. Teachers need to know about their pupils' progress and difficulties with learning so that they can adapt their own work to meet pupils' needs -- needs that are often unpredictable and that vary from one pupil to another. Teachers can find out what they need to know in a variety of ways, including observation and discussion in the classroom and the reading of pupils' written work.

We use the general term assessment to refer to all those activities undertaken by teachers -- and by their students in assessing themselves -- that provide information to be used as feedback to modify teaching and learning activities. Such assessment becomes formative assessment when the evidence is actually used to adapt the teaching to meet student needs.²

There is nothing new about any of this. All teachers make assessments in every class they teach. But there are three important questions about this process that we seek to answer:

- Is there evidence that improving formative assessment raises standards?
- Is there evidence that there is room for improvement?
- Is there evidence about how to improve formative assessment?

The conclusion we have reached from our research review is that the answer to each of the three questions above is clearly yes.

Does Improving Formative Assessment Raise Standards?

A research review published in 1986, concentrating primarily on classroom assessment work for children with mild handicaps, surveyed a large number of innovations, from which 23 were selected.⁴ Those chosen satisfied the condition that quantitative evidence of learning gains was obtained, both for those involved in the innovation and for a similar group not so involved. Since then, many more papers have been published describing similarly careful quantitative experiments. Our own review has selected at least 20 more studies. (The number depends on how rigorous a set of selection criteria are applied.) All these studies show that innovations that include strengthening the practice of formative assessment produce significant and often substantial learning gains. These studies range over age groups from 5-year-olds to university undergraduates, across several school subjects, and over several countries.

For research purposes, learning gains of this type are measured by comparing the average improvements in the test scores of pupils involved in an innovation with the range of scores that are found for typical groups of pupils on these same tests. The ratio of the former divided by the latter is known as the effect size. Typical effect sizes of the formative assessment experiments were between 0.4 and 0.7. These effect sizes are larger than most of those found for educational interventions. The following examples illustrate some practical consequences of such large gains.

- An effect size of 0.4 would mean that the average pupil involved in an innovation would record the same achievement as a pupil in the top 35% of those not so involved.
- An effect size gain of 0.7 in the recent international comparative studies in mathematics⁵ would have raised the score of a nation in the middle of the pack of 41 countries (e.g., the U.S.) to one of the top five.

Thus it seems clear that very significant learning gains lie within our grasp. The fact that such gains have been achieved by a variety of methods that have, as a common feature, enhanced formative assessment suggests that this feature accounts, at least in part, for the successes. However, it does not follow that it would be an easy matter to achieve such gains on a wide scale in normal classrooms. Many of the reports we have studied raise a number of other issues.

- All such work involves new ways to enhance feedback between those taught and the teacher, ways that will require significant changes in classroom practice.
- Underlying the various approaches are assumptions about what makes for effective learning -- in particular the assumption that students have to be actively involved.
- For assessment to function formatively, the results have to be used to adjust teaching and learning; thus a significant aspect of any program will be the ways in which teachers make these adjustments.
- The ways in which assessment can affect the motivation and self-esteem of pupils and the benefits of engaging pupils in self-assessment deserve careful attention.