

Schools have to produce an overview of progress in their students' learning, in the form of a summative assessment, at least once a year. National testing policies may or may not provide helpful support for this work, but they cannot supplant the responsibility of schools to provide regular summary guidance about the progress of each student's learning. Some see this task as an unfortunate necessity, and many experience a tension, between the formative practices which have been shown to enhance students' learning, and their own summative assessment requirements.

This booklet builds on work that the King's College London team have done in collaboration with teachers in three schools to develop ways of reviewing and improving their summative assessment methods. That collaboration developed a practical approach to this task; one which has shown that such development can forge consistent and positive policies which can establish mutually supportive relationships between formative and summative assessments. These relationships are part of an approach which treats assessment as an intrinsic part of effective teaching and learning.

In addition to reporting these positive findings, this booklet also shows how

they are reinforced by a survey of reports of research and development projects from six other countries. These also show that work aimed at enhancing the quality and status of teachers' summative judgements used methods similar to those we developed, and produced similar positive effects on their teaching.

A finding common to those projects and to our own is that collaboration between teachers is essential to their professional development in assessment, and that work which such collaboration produces can both enrich the quality of their teaching and enhance teachers' confidence in their own assessment results.

## ACKNOWLEDGEMENTS

Most of the guidance offered in this booklet derives from the collaboration of its authors with teachers in three schools in Oxfordshire. In this collaboration, we and the teachers explored how the principles of summative assessment and the ideas reported in the wide range of international literature about this topic, could be implemented within the everyday context of the schools. We can claim that the ideas that are set out here are practicable and rewarding because they emerged

from the practical experience and the reflective insights of the teachers who worked with us. This work owes a great deal to these teachers, and to the support of their work by their schools and by the local authority. We also acknowledge the contribution of the Nuffield Foundation: their financial support was crucial in enabling us to carry out this work.

## 1. WHAT DOES THIS BOOKLET OFFER?

Previous booklets in this Black Box series have focused on teachers' formative assessment, indicating ways in which assessment can serve, at the heart of the day-to-day activities in the classroom, to improve students' learning. Such activities are central to every teacher's work, and many have welcomed ideas about ways to make their assessment more effective.

It would be surprising if all ideas about improving testing were to be welcomed with the same enthusiasm by everyone. While teachers have readily engaged with the guiding principles of formative assessment, we have found that they are less willing to look at changing their summative assessment practices. One major reason for this is the stress and pressures caused by externally

imposed tests. School managements and parents require 'success' so that many teachers feel compelled to 'teach to the test' even though they admit that by so doing they are betraying their own values. A more fundamental effect of the political commitment to external accountability systems is that it suggests to teachers that many politicians, parents and even some of their own school's management, believe teachers' assessments of their own students are flawed and cannot be trusted. Is this scepticism justified?

In the UK, national inspectorates' reports have repeatedly identified assessment as the weakest aspect of the work of schools. Moreover, other countries have found that when externally imposed tests are abandoned, many teachers continue to use the same types of tests as the sole basis for their summative assessments. So, while teachers may have disliked the external tests, many were not confident that they could do better.

Does this matter? A concern frequently expressed by teachers who have been committed to improving assessment for learning is that at some point they have to change gear and teach to the test: the summative and the formative functions are seen at worst as irreconcilable, at best as uneasy bedfellows. A more supportive

relationship would clearly be helpful for teachers and for their students.

But there are other reasons why teachers' lack of confidence in their ability to formulate effective tests matters. Teachers' assessment of their students is not just for external accountability, year-on year results are important for the school management's assessment of individual teachers' work. Furthermore, advice to students and their parents about progress and about future choices among subjects, and communication to future teachers of the same students all depend on the summative assessments of current and past teachers. A notable example is the transition of students between any two stages, whether from primary to secondary, or between upper secondary stages.

More generally, policy-makers' distrust of teachers' abilities to assess should ideally be overcome in order for students and their parents to trust teachers in their turn. Such distrust can only be challenged if teachers and schools are confident that the skills, practices and policies underpinning the summative assessments they carry out can be trusted as fair to every student and comparable in standard within and between schools.

In principle, a school's own assessments should be more helpful

to students than their marks on an external test. Indeed, since the teacher has numerous opportunities over a year's work with a class to observe each student's achievements and problems, it seems ridiculous that the evidence of a few hours' work in the artificial conditions of the formal written test be preferred to any evidence that the teacher can produce. Since teachers' are not subject to the same constraints of time and cost as external agencies, they ought to be able to take account of students' achievements in a range of contexts far broader than those of the formal written test. Thus, teachers' assessments should, in principle, be better able to meet the criteria of reliability, validity and comparability by which such agencies are themselves judged.

The purpose of this booklet is to set out ways in which the summative assessments created by teachers and their schools can be improved so that they comfortably meet the criteria upon which all such assessment should be judged. Its ideas and recommendations are based on two main sources. The first is our experience of, and findings from, a two-year development study with three secondary schools in England in which approximately 20 teachers of English and of mathematics worked in collaboration to improve the quality of their summative assessments and to

establish that the results were fair and comparable across the three schools (Black et al. 2011).

The second source is the findings of published studies, notably from USA, Canada, Australia, New Zealand, Scotland and Wales. These studies show that the issues outlined above have been a cause for concern in many different state systems. Moreover, the different situations, and the innovations introduced in pursuit of reform described in these studies, help to broaden understanding of the problems involved and may assist in the development of ways to tackle them.

This booklet sets out its analysis and recommendations in the following sequence. *Section 2* gives a brief account of the roles of two main functions of assessment, *for* learning and *of* learning, within the whole enterprise of teaching and learning, and also outlines the five stages of any teaching activity. *Section 3* then considers the quality of assessments, exploring validity, reliability and comparability and the challenges that can undermine them.

These two brief sections serve to set up the background for *Section 4*, which describes a seven-step action plan for summative assessment, outlining the

methods developed in our work with our group of teachers in England. It shows how teachers found that their active participation improved both the quality of their summative assessments and their confidence in the results that they produced. Our approach was not to seek to impose a new system but rather to look at the methods that teachers already used in their classrooms and then help them develop these in a more rigorous and dependable fashion. Our hope is that this booklet will help other teachers in a similar way. This help may be further enriched by the evidence and ideas in *Section 5*, which reflects on, and adds to, these findings in the light of further lessons that can be drawn from work from other countries, so leading to the summary of our main conclusions in *Section 6*.

## 2. ASSESSMENT INSIDE THE BLACK BOX

Both formative and summative assessment are essential components of teaching and learning. In general terms, assessment is simply the production and interpretation of evidence of achievement. If this evidence is used to guide the next steps in progress, it is for learning, or formative; if it is used to sum up,

judge, make decisions about, progress so far, it is of learning, or summative. The results of any assessment might serve both categories.

A unified view of the role of assessment in pedagogy can be presented in terms of the five stages involved in any teaching activity, as follows:

### **A. FORMULATING AIMS**

This is the stage of strategic decision. All that follows should relate to a clear formulation of the learning aims.

### **B. PLANNING ACTIVITIES**

The aims are to be achieved by choosing, adapting, or inventing activities which will engage students and thereby elicit responses from them which help to clarify and then extend their understanding.

### **C. IMPLEMENTATION**

The way in which a plan is implemented in the classroom is crucial. What is needed is formative interaction which stimulates and builds on the students' contributions. Thus, implementation of a plan based on stated aims is the core activity of assessment for learning.

### **D. REVIEW**

At the end of any learning episode there should be a review, to check before moving on. The assessment used at this stage may be designed to

be summative, but its results can also be for learning. An example is using peer marking to help all students to develop understanding of the criteria required to meet the aims. It may also help the teacher to identify a need to revisit some issues with the class as a whole.

### **E. SUMMING UP**

This is a more formal version of the Review stage above; here the results may be used to make decisions about a student's future work or career, to report progress to other teachers, school management teams and parents, and to report the overall achievements more widely to satisfy the need for accountability.

These five stages can interact in helpful ways. The experience of classroom implementation in stage C may lead to changes in the activities planned in stage B; the findings of a review in stage D could prompt a revision of the work in either, or both, of stages C and B. Stage E might go further, as the broader overview that should be characteristic of, say, an end-of-year assessment might indicate ways in which all four preceding stages should be reconsidered.

In addition to such *feed-back* of each stage to its predecessors, there should also be *feed-forward*. This second idea