

## Chapter 2

---

# Assessment for Learning in the Classroom

*Paul Black and Dylan Wiliam*

### Introduction

Assessment in education must, first and foremost, serve the purpose of supporting learning. So it is fitting to start a study of assessment with an exploration of the meaning and practices of assessment that serve this purpose most directly. This chapter is the story of a development that started with a review of what research had to say about formative assessment. The background to this review, and the main features of its findings, are first described. Its results led to development work with teachers to explore how ideas taken from the research could be turned into practice. A description of this work is followed by reflections on outcomes and implications. Finally, we will the dissemination of the project's findings and its wider impact.

### The research review

#### *The background*

Studies over many years have shown that formative assessment is an important aspect of teachers' classroom work and that attention to improving its practice can enhance the learners' achievements. Harry Black, a researcher in Scotland, who was unique at the time in working with teachers to develop formative assessment, introduced his account of the subject by pointing out that formative assessment has always been part of the practice of teachers, quoting in evidence a

letter written by the Principal of Greenwich Hospital School (quoted in Chadwick, 1864), and calling attention to its neglect in the following trenchant terms:

Consider the amount of time, energy and money spent by both individual teachers, and schools in general, on setting and marking continuous assessment tests, end of session examinations and mock O' levels. Reflect on the money spent by examination boards and the number of assessment specialists employed by them. Read, if you can find a sabbatical term, the literature on the technology of assessment for reporting and certification. Compare these in turn with the complete lack of support normally given to teachers in devising and applying procedures to pinpoint their students' learning problems, with the virtual absence of outside agencies to develop formative assessment instruments and procedures, and the limited literature on the topic. (Black, 1986: 7)

Linn, writing three years later, made a different and prophetic point about what might be involved:

the design of tests useful for the instructional decisions made in the classroom requires an integration of testing and instruction. It also requires a clear conception of the curriculum, the goals, and the process of instruction. And it requires a theory of instruction and learning and a much better understanding of the cognitive processes of learners. (Linn, 1989: 5)

These extracts should not be misunderstood: it is clear from Harry Black's work that his terms 'procedures' and 'instruments' were not references to conventional summative tests, and it should also be clear that Linn's 'tests useful for ... instructional decisions' was not a reference to such tests either. However, despite such insights in the writing of several authors, formative assessment was not regarded as a more than marginal component by many involved in public debates on education, even when the report of the government's Task Group on Assessment and Testing (TGAT; National Curriculum Task Group, 1988) made teachers' formative assessment a central plank for its proposals.

Yet there was accumulating in the research literature on formative assessment practices a formidable body of evidence that could support claims for its importance. Early reviews by Natriello (1987), Crooks (1988) and Black (1993) drew attention to this evidence, but these were neither sufficiently comprehensive in scope, nor targeted

directly at making the argument that formative assessment was a powerful way to raise standards. The Assessment Reform Group, however, whose main concern was consideration of research evidence as a basis for formation of assessment policy, judged that further exploration of formative assessment was essential and in 1996 obtained funding from the Nuffield Foundation to support a review of research. The Group then invited the two of us to carry out this review, and because of our long-standing interest in formative assessment, we were happy to agree to undertake the task.

### *The review*

Our survey of the research literature involved checking through many books, through the issues of over 160 journals published between 1988 and 1997, and studying earlier reviews of research (Crooks, 1988; Natriello, 1987). This process yielded about 580 articles or chapters to study. Out of this we prepared a lengthy review that cited 250 of these sources. The review was published (Black and Wiliam, 1998a) together with comments on our work by experts from five different countries. As this work progressed, we developed a new view of the issues relevant to the field; a view that we set out in six sections.

A first section surveyed the evidence. We looked for studies that showed quantitative evidence of learning gains by comparing data for an experimental group with similar data from a control group. We reported on about 30 such studies, all of which showed that innovations which included strengthening the practice of formative assessment produced significant, and often substantial, learning gains. They ranged over ages (from 5-year-olds to university undergraduates), across several school subjects, and over several countries.

The fact that such gains had been achieved by a variety of methods that had, as a common feature, enhanced formative assessment indicated that it is this feature that accounted, at least in part, for the successes. However, it did not follow that it would be an easy matter to achieve such gains on a wide scale in normal classrooms, in part because the research reports lacked enough detail about the practical use of the methods, detail that would be essential if replication were envisaged. More significantly, successful implementation of such innovations is dependent on the social and educational context of their development, so that they cannot merely be implemented in the same way if they are to be successful in a different context.

A second section covered research into current practices of teachers. The picture that emerged was that, for the vast majority of teachers, formative assessment was not a well-developed aspect of practice.

In relation to *effective learning* it seemed that teachers' questions and tests encouraged rote and superficial learning, even where teachers said that they wanted to develop understanding. There was also evidence of the *negative impact* of a focus on comparing students with one another, so emphasizing competition rather than personal improvement. Furthermore, teachers' feedback to students often seemed to serve social and managerial functions, often at the expense of the learning functions.

A third section focused on research into the involvement of students in formative assessment. Students' beliefs about the goals of learning, about the risks involved in responding in various ways, and about what learning work should be like, were all shown to affect their motivation. Other research explored the different ways in which positive action could be taken, covering such topics as study methods, study skills, and peer- and self-assessment.

A fourth section looked at ideas that could be gleaned from the research about strategies that might be productive for teachers. One feature that emerged was the potential of the learning task, as designed by a teacher, for exploring students' learning. Another was the importance of the classroom discourse, as steered by teachers' questions and by their handling of students' responses.

A fifth section shifted attention to research into comprehensive systems of teaching and learning in which formative assessment played a part. One example was mastery learning programmes. In these it was notable that students were given feedback on their current achievement against some expected level of achievement (i.e. the 'mastery' level); that such feedback was given promptly; and that students were given the opportunity to discuss with their peers how to remedy any weaknesses.

A sixth section explored in more detail the literature on feedback. The review of empirical evidence by Kluger and DeNisi (1996) showed that feedback can have positive effects only if the feedback is used as a guide to improvement, whilst the conceptual analysis of the concept of feedback by Ramaprasad (1983) and the development of this by Sadler (1989) emphasized that learners must understand both the 'reference level' (i.e. the goal of their learning) and the actual level of their understanding. Another important message came from the research on attribution theory (for example by Vispoel and Austin [1995] and by Dweck [2000]) that teachers must aim to inculcate in their students the idea that success is due to internal, unstable, specific factors such as effort, rather than on stable general factors such as ability (internal) or whether one is positively regarded by the teacher (external).

Four key themes emerged from the research we reviewed. First, formative work involves new ways to enhance feedback between those taught and the teacher, ways which require new modes of pedagogy and significant changes in classroom practice. Second, underlying the various approaches are assumptions about what makes for effective learning – in particular that students have to be actively involved. Third, for assessment to function formatively, the results have to be used to adjust teaching and learning – so a significant aspect of any programme will be the ways in which teachers do this. Fourth, the ways in which assessment can affect the motivation and self-esteem of students, and the benefits of engaging students in self-assessment, both deserve careful attention.

The structure of the six sections outlined above did not emerge automatically: as our work progressed, so we had to find ways of organizing the field, and making new conceptual links in order to be able to combine the various findings into as coherent a picture as possible. We believe that our review generated a momentum for work in this field by providing a new framework that would be difficult to create in any other way.

## Moving into action

### *Setting up a project*

Given that our review had shown that innovations in formative assessment could raise standards of student achievement substantially, it was natural to think about ways to help schools secure these benefits. However, even if a recipe for practice could have been derived from the variety of research studies, our own experience of teachers' professional development had taught us that the implementation of innovative practices in classrooms could not be a straightforward matter of proclaiming a recipe for teachers to follow. We believed that new ideas about teaching and learning could only be made to work in particular contexts, in our case that of teachers in (initially) secondary schools in the UK, if teachers were able to transform or 'morph' them (Ginsburg, 2001) and so create new practical knowledge relevant to their work.

We obtained funding from the UK's Nuffield Foundation (and later also from the National Science Foundation in the USA), for a two-year development project. To find schools and teachers to work with, we talked with assessment specialists from two local education authorities (usually called school districts in the USA, and local

authorities, or just LAs in England). The two were chosen because we knew that they would understand and support our aims, and visits to their respective districts (Oxfordshire and Medway) could reasonably be managed in a day.

In each local authority, three schools teaching students from 11 to 18 years of age were then chosen by the LA specialists. The initial choice was made by the specialists, although we had requested that schools that were either in serious difficulties or unusually successful would not be chosen. The schools agreed to collaborate with us and each school identified two science and two mathematics teachers. In the second year of the project we added two teachers of English, from each of the same schools, and one additional mathematics and science teacher, so that in all 48 teachers were involved. The LA specialists were involved with the work throughout. The project was called the King's College Medway Oxford Formative Assessment Project (KMOFAP) to highlight our close collaboration with these partners (Black and Wiliam, 2003).

Because the teachers within the schools were chosen by the schools themselves (sometimes with advice from the LA specialists), the teachers varied considerably in their experience, qualifications and expertise. Some were newly qualified, others were heads of the subject departments, one was close to retirement, and one apparently was identified because the head teacher of the school thought she 'needed some INSET'. However the majority were experienced and well qualified. Before the start of the project's work, we and the LA specialist visited each school in order to explain the aims and the requirements to the head teacher. We chose to work with secondary mathematics and science because we were specialists in those two subjects at this level and believed that the nature of the subject matter was important. English teachers were brought in when a colleague specializing in English was able to join the team.

We advised the teachers to focus on a few, and eventually just one, focal class to try out the innovations that they chose to work at, with a caveat that they might do well to avoid those age groups (13–14 and 15–16-year-olds), for which statutory national tests might inhibit their freedom to experiment. In the event, some ignored this advice, so that the classes involved ranged over ages 11 to 16. Whilst support from each school's senior management was promised in principle, it varied in practice; moreover, within the school subject faculty or department, some had stronger support from subject colleagues than others, and in fact the collegial support that would be essential in an endeavour of this kind was largely provided by the meetings, once every five weeks, when the project teachers all spent a day together with the staff at King's. There was evidence of interest and support

from other school colleagues – several productive ideas were injected into the group from this type of source, and it was soon clear that the ideas in the project were influencing teachers more widely, to the extent that in some cases it was difficult to find suitable ‘control’ classes for comparison of their test performance with those of pupils in the focal classes of the project.

### *The practices developed*

These practices will be described here under four headings: oral feedback in *classroom questioning* (more recently re-labelled as *dialogue*), *feedback through marking*, *peer- and self-assessment*, and the *formative use of summative tests*. The account given will be brief – more detailed accounts have been published elsewhere (Black et al., 2003). These practices were defined and developed in the course of the project, the process being one in which we drew, from the research findings, a variety of ideas for which there was evidence of potential value, and then teachers selected from these and developed them in their own ways. The four themes discussed below were an outcome of the project, for while they were related to our inputs, we could not have predicted at the outset that a set of themes would emerge in the way that they did.

For *classroom dialogue* the aim was to improve the interactive feedback that is central to formative assessment. After hearing an account of research on wait-time (e.g., Rowe, 1974) teachers were motivated to allow a longer time after asking a question so that students would have time to think out responses, and so that all could be expected to become actively involved in question and answer discussions, and to make longer replies. Increased participation of students also required that all answers, right or wrong, be taken seriously, the aim being to develop thoughtful improvement rather to evoke the expected answers. A consequence of such changes was that teachers learnt more about the pre-knowledge of their students, and about any gaps and misconceptions in that knowledge, so that their next ‘moves’ could address the learners’ real needs.

As they tried to develop this approach, teachers realized that more effort had to be spent in framing questions that would evoke, and so help to explore, critical indicators of students’ understanding. They also had to listen carefully to students and then formulate meaningful responses and challenges that would help them to extend that understanding.

The task of developing an interactive style of classroom dialogue required a radical change in teaching style from many teachers, one that they found challenging. Some were well over a year into the

project before such change was achieved. Subsequent work with other schools has shown that it is this aspect of formative work that teachers are least likely to implement successfully.

To address *feedback through marking*, teachers were first given an account of research studies that established that, while students' learning can be advanced by feedback through comments, the giving of marks or grades has a negative effect because students ignore comments when marks are also given (Butler, 1988). These results surprised and worried the teachers, because of concern about the effect of returning students' work with comments but no marks. However, potential conflicts with school policy were resolved as the teachers discovered that providing comments rather than grades gave both students and their parents advice on how to improve. It also set up a new focus on the issue of how to move learning forward rather than on trying to interpret a mark or grade. To make the most of the learning opportunity created by feedback on written work, procedures that required students to follow up comments had to be planned as part of the overall learning process.

One consequence of this change was that teachers had to think more carefully in framing comments on written work in order to give each student guidance on how to improve. As the skills of formulating and using such feedback were developed, it became clearer that the quality of the tasks set for written homework or class work was critical: as for oral questions, tasks had to be designed to encourage students to develop and express key features of their understanding.

For *peer- and self-assessment*, the starting point was Sadler's (1989) argument that self-assessment is essential to learning because students can only achieve a learning goal if they understand that goal and can assess what they need to do to reach it. Thus the criteria for evaluating any learning achievements must be made transparent to students to enable them to have a clear overview both of the aims of their work and of what it means to complete it successfully. In so far as they do so they begin to develop an overview of that work so that they can manage and control it: in other words, they develop their capacity for meta-cognitive thinking. A notable example of the success of such work is the research of White and Frederiksen (1998).

In practice, peer-assessment turned out to be an important stimulus to self-assessment. It is uniquely valuable because the interchange will be in language that students themselves would naturally use, because students learn by taking the roles of teachers and examiners of others (Sadler, 1998), and because students appear to find it easier to make sense of criteria for their work if they examine other students' work alongside their own. A typical exercise would be on the

marking of homework. Students were asked to label their work with 'traffic lights' as an indicator of their confidence in their learning (i.e. using red or amber if they were totally or partially unsure of their success, and green where they were confident). Then those who had used amber or green would work in mixed groups to appraise and help with one another's work, whilst the teacher would pay special attention to those who had chosen red.

Teachers developed three ways of making *formative use of summative tests*. One way was to ask students, in preparation for a test, to 'traffic light' a list of key words or of the topics on which the test would be set, an exercise which would stimulate them to reflect on where they felt their learning was secure and where they needed to concentrate their efforts. One reason for doing this was that teachers had realized that many students had never thought about developing a strategy for preparing for a test such as formulating a strategic appraisal of their learning.

A second way was to mark one another's test papers in peer groups, in the way outlined above for the marking of homework. This could be particularly challenging when they were expected to invent their own marking rubric, for to do this they had to think about the purpose of a question and about the criteria of quality to apply to responses. After peer marking, teachers could reserve their time for discussion of the questions that gave particular difficulty.

A further idea was introduced from research studies that have shown that students trained to prepare for examinations by generating and then answering their own questions out-performed comparable groups who prepared in conventional ways (Foos et al., 1994; King, 1992). Preparation of test questions calls for, and so develops, an overview of the topic.

The teachers' work on summative assessments challenged our expectations that, for the context in which they worked, formative and summative assessments are so different in their purpose that they have to be kept apart. The finding that emerged was quite different – that summative tests should be, and should be seen to be, a positive part of, and therefore integrated into, the learning process. If they could be actively involved in the test process, students might see that they can be beneficiaries rather than victims of testing, because tests can help them improve their learning. However, this synergy could not be achieved in the case of high-stakes tests set and marked externally; for these, as currently designed and administered, formative use would not be possible; it can be achieved in the case of summative tests designed and used for internal use in a school (see Black et al., 2010)

## Reflections on the outcome

It was clear that the new ideas that had emerged between the teachers and ourselves involved far more than the mere addition of a few tactical tricks. Some reflection was needed to tease out the more fundamental issues that seemed to be raised.

### *A focus on learning*

One of the most surprising things that happened during the early project meetings was that the participating teachers asked us to run a session on learning theories. In retrospect, perhaps, we should not have been so surprised. Whilst teachers could work out after the event whether or not any feedback had had the desired effect, what they needed was to be able to give their students feedback that they knew in advance was going to be useful. To do that they needed to build up models of how students learn.

As a result, the teachers came to take greater care in selecting tasks, questions, and other prompts, to ensure that the responses made by students actually 'put on the table' the ideas that they bring to a learning task. The key to effective learning is to then find ways to help students restructure their knowledge to build in new and more powerful ideas. In the KMOFAP classrooms, as the teachers came to listen more attentively to the students' responses, they began to appreciate more fully that learning is not a process of passive reception of knowledge, but one in which the learners are active in creating their own understandings. These ideas reflect some of the main principles of the constructivist view of learning – to start where the student is and to involve the students actively in the process, and to understand that because students are active in the construction of their own knowledge, what they construct may be very different from what the teacher intended.

Students also changed, coming to understand what counted as good work through a focus on the criteria and on their exemplification. Sometimes this was done through focused whole-class discussion around a particular example; at other times it was achieved through students using criteria to assess the work of their peers. The activities, by encouraging students to review their work in the light of the goals and criteria, were helping them to develop meta-cognitive approaches to learning.

Finally, the involvement of students both in whole-class dialogue and in peer-group discussions, as part of a broader shift in the classroom culture (to which all four activities contributed), created a richer community of learners where the social learning of students would become more salient and effective.

### *A learning environment and changes of role*

Reflection on the experiences described above led to more profound thinking by participants about their role as teachers and about the need to 'engineer' learning environments in order to involve students more actively in learning tasks. The emphasis had to be on the students doing more of the thinking and making that thinking public. As one teacher said:

There was a definite transition at some point, from focusing on what I was putting into the process, to what the students were contributing. It became obvious that one way to make a significant sustainable change was to get the students doing more of the thinking. I then began to search for ways to make the learning process more transparent to the students. Indeed, I now spend my time looking for ways to get students to take responsibility for their learning and at the same time making the learning more collaborative. Tom, Riverside School (Black et al., 2002)

This teacher had changed his role from being a presenter of content to being a leader of an exploration and development of ideas in which all students were involved. One of the striking features of the project was the way in which, in the early stages, many spoke about the new approach as 'scary', because they felt that they were losing control of their classes. Towards the end of the project, they described this same process not as a loss of control, but as one of sharing responsibility for the class's learning with the class – exactly the same process, but viewed from two very different perspectives.

The learning environment envisaged requires a classroom culture that may well be unfamiliar and disconcerting for both teachers and students. The effect of the innovations implemented by our teachers was to change the rules, usually implicit, that govern the behaviours that are expected and seen as legitimate by teachers and by students. As Perrenoud (1991: 92) put it: 'Every teacher who wants to practise formative assessment must reconstruct the teaching contract so as to counteract the habits acquired by his pupils'.

For the students, they have to change from behaving as passive recipients of the knowledge offered to becoming active learners who could take responsibility for their own learning. These students became more aware of when they were learning, and when they were not. One class, who were subsequently taught by a teacher not emphasizing assessment for learning, surprised that teacher by complaining: 'Look, we've told you we don't understand this. Why are you going on to the next topic?'

Another way of thinking about what happened in the project is that the role expectations – that is, what teachers and students thought that being a teacher or being a student required you to do – had been altered. Whilst it can seem daunting to undertake such changes, they do not have to happen suddenly. Changes with the KMOFAP teachers came slowly, typically over two years rather than one, and steadily, as experience developed and confidence grew in the use of the various strategies for enriching feedback and interaction.

A collection of individual and group discussion data near the end of the project did expose one unresolved problem – the tension between the formative approach and summative demands. Some, but not all, teachers were confident that the new work would yield better test result than ‘teaching to the test’. However, for their in-school summative tests, many felt impelled to use questions from the key stage 3 and GCSE tests despite doubts about the validity of these in relation to the improved pupil learning achieved in the project. The general picture was that, despite developing the formative use of their summative tests, teachers felt that they could not reconcile the external test and accountability pressures with their investment in improved formative assessment.

## Research and practice

### *Explaining success: the focus of the project*

We were surprised that the project was so successful in promoting quite radical changes in the practices of almost all of the teachers involved, and wondered whether lessons could be learned from it about the notoriously difficult problem of turning research into practice. One relevant factor is that the ideas that the project set before the teachers had an intrinsic acceptability to them. We were talking about improving learning in the classroom, which was central to their professional identities, as opposed to bureaucratic measures such as predicting test levels. One feature of our review was that most of it was concerned with such issues as students’ perceptions, peer- and self-assessment, and the role of feedback in a pedagogy focused on learning. Thus it helped to take the emphasis in formative assessment studies away from *systems*, with its emphasis on the formative-summative interface, and re-locate it on classroom *processes*. Acceptability was also enhanced by our policy of emphasizing that it was up to each teacher to make his or her own choice between the different formative practices; so teachers developed their own personal portfolios,

adding to or dropping components as experience and the experiences of their colleagues led them to change.

Linked to the previous factor is that in our choice to concentrate on the classroom processes, we had decided to live with the external constraints operating at the formative–summative interface: the legislated attempts to change the *system*, in the 1980s and 1990s in England, were set aside. Whilst it might have been merely prudent to not try to tilt at windmills, the more fundamental strength was that it was at the level chosen – learning processes in classrooms – that formative work stakes its claim for attention. Furthermore, given that any change has to work out in teachers' practical action, this is where reform should always have started.

Another factor that appears to have been important is the credibility that we brought as researchers to the process. In their project diaries, several of the teachers commented that it was our espousal of these ideas, as much as the ideas themselves, that persuaded them to engage with the project. Part of that credibility is that we chose to work with teachers in the three subjects, English, mathematics and science, when, in each of these, one or two members of the team had expertise and reputations in the subject community. Thus, when specific issues, such as 'Is this an appropriate question for exploring students' ideas about the concept of photosynthesis?' arose, we could discuss them seriously.

### *Explaining success: the process strategy*

The way in which teachers were involved was also important. They all met with the researchers for a whole day every five weeks, over a period of two years. In addition, two researchers were able to visit the schools, observe the teachers in their classrooms, give them feedback, collect interview data on their perceptions, and elicit ideas about issues for discussion in the whole-day meetings. The detailed reports of our findings (Black et al., 2002, 2003) are based on records of these meetings, on the observations and records of visits to classrooms by the King's team, on interviews with and writing by the teachers themselves, on feedback from the LA specialists who held their own discussions with their teachers, and on a few discussions with student groups. As the project developed, the King's team played a smaller part as the teachers took over the agenda and used the opportunity for their own peer learning.

In our development model, we attended to both the *content* and the *process* of teacher development (Reeves et al., 2001). We attended to the *process* of professional development through an acknowledgment that teachers need time, freedom, and support from colleagues,

in order to reflect critically upon and to develop their practice (Lee and Wiliam 2000), whilst offering also practical strategies and techniques about how to begin the process. By themselves, however, these are not enough. Teachers also need concrete ideas about the directions in which they can productively take their practice, and thus there is a need for work on the professional development of teachers to pay specific attention to subject-specific dimensions of teacher learning (Wilson and Berne, 1999).

One of the key assumptions of the project was that if the promise of formative assessment was to be realized, a research design in which teachers are asked to test out, and perhaps modify, a scheme worked out for them by researchers would not be appropriate. We presented them with a collection of ideas culled from research findings rather than with a structured scheme. We argued that a process of supported development was an essential next step. In such a process, the teachers in their classrooms had to work out the answers to many of the practical questions that the research evidence that we presented could not answer. The issues had to be reformulated in collaboration with them, where possible in relation to fundamental insights, and certainly in terms that could make sense to their peers in ordinary classrooms.

The key feature of the INSET sessions was the development of action plans. Since we were aware from other studies that effective implementation of formative assessment requires teachers to renegotiate the 'learning contract' that they had evolved with their students (Brousseau, 1984; Perrenoud, 1991), we decided that implementing formative assessment would best be done at the beginning of a new school year. For the first six months of the project (January 1999 to July 1999), therefore, we encouraged the teachers to experiment with some of the strategies and techniques suggested by the research, such as rich questioning, comment-only marking, sharing criteria with learners, and student peer- and self-assessment. Each teacher was then asked to draw up an action plan of the practices they wished to develop and to identify a single focal class with whom these strategies would be introduced at the start of the new school year in September 1999. Details of these plans can be found in Black et al. (2003). As the teachers explored the relevance of formative assessment for their own practice, they transformed ideas from the research and from other teachers into new ideas, strategies and techniques, and these were in turn communicated to teachers, creating a 'snowball' effect. As we have introduced these ideas to more and more teachers outside the project, we have become better at communicating the key ideas (see Chapter 3 for further exploration of this issue).

Through our work with teachers, we have come to understand more clearly how the task of applying research to practice is much

more than a simple process of ‘translating’ the findings of researchers into the classroom. The teachers in our project were engaged in a process of knowledge creation, albeit of a distinct kind and possibly relevant only in the settings in which they work (Hargreaves, 1999). We stressed this feature of our approach with the teachers right from the outset of the project. We discovered later that some of them did not, at that stage, believe us: they thought that we knew exactly what we wanted them to do but wanted them to discover it for themselves. As they came to know us better, they realized that, at the level of everyday classroom practice, we really did not know what to do. The arguments in this section are addressed only to the specific question with which it started – why did this project work – with the intent of thereby illuminating the vexed issues of the relationship between research and practice. They cannot claim to address the question of whether an innovation with similar aims would succeed in different circumstances. Any attempt to answer such a question would have to relate the context and particular features of our work to the context and features of any new situation, bearing in mind that any such innovation will start from where our work finished and not from where we started.

## Dissemination and impact

### *Publicity*

Publicity designed to make a case for formative assessment started, alongside the publication of the research review, in 1998. Although we tried to adhere closely to the traditional standards of scholarship in the social sciences when conducting and writing our review, we did not do so when exploring the policy implications in a booklet, entitled *Inside the Black Box* (Black and Wiliam, 1998b) that we published, and publicized widely, alongside the academic review. This raised a great deal of interest and created some momentum for our project and for subsequent dissemination. While the standards of evidence we adopted in conducting the review might be characterized as those of ‘academic rationality’, the standard for *Inside the Black Box* was much closer to that of ‘reasonableness’ as advocated by Stephen Toulmin for social inquiry (Toulmin, 2001). In some respects, *Inside the Black Box* represented our opinions and prejudices as much as anything else, although we would like to think that these are supported by evidence, and are consistent with the 50 years of experience in this field that we had between us. It is also important to note that the success of *Inside the Black Box* – it has to

date sold about 50,000 copies – has been as much due to its rhetorical force as to its basis in evidence, whilst the version published in a USA teacher journal has been the most frequently quoted article in that journal. This would make many academics uneasy – for it appears to blur the line between fact and value, but as Flyvbjerg (2001) argues, social enquiry has failed precisely because it has focused on analytic rationality rather than value-rationality (see also Wiliam, 2003).

The quantitative evidence that formative assessment does raise standards of achievement was a powerful motivator for the teachers at the start of the project. One aspect of the KMOFAP project was that the King's team worked with each teacher to collect data on the gains in test performance of the students involved in the innovation, and comparable data for similar classes who were not involved (Wiliam et al., 2004). The project did not introduce any tests of its own but rather relied on achievement data used from the tests that the schools used for all students, whether or not they were involved in the project. The analysis of these data showed an overall and significant gain in achievement outcomes. Thus the evidence from the research review can now be supplemented by evidence of enhanced performance on the UK national and on schools' own examinations. This evidence was incorporated, with an account of the practical lessons learnt in the KMOFAP project, in a second small booklet, *Working Inside the Black Box* (Black et al., 2002), which has also been widely successful with over 50,000 copies sold to date, whilst a detailed account of the project's work (Black et al., 2003) has also been very well received. Further booklets on specific aspects of formative assessment have also been produced<sup>1</sup> and other publicity for formative assessment, further research results and practical advice, notably from the Assessment Reform Group (ARG: 1999, 2001, 2002; Mansell et al., 2009) have added to the impact.

### *Dissemination*

Following this project, members of the King's team have responded to numerous invitations to talk to other groups: over three years they have made over 500 such contributions. These have ranged across all subjects, across primary, secondary and post-compulsory phases. In addition, there has been sustained work with four groups of primary schools. The King's team has also been involved as advisers to large-scale development ventures, in several local government districts in the UK, and with education ministries in Scotland and in Jersey.

The Education Department of the Scottish Executive, which has full legislative powers in education in Scotland, has taken up the work

as one of several strands of its Assessment is for Learning Development Programme. This project, entitled Support for Professional Practice in Formative Assessment, involved four groups of eight or nine schools, including both secondary and primary. They were supported by one development officer and staff from two university faculties, and also by contributions from the King's project staff. The work started in May 2002, and an evaluation project, conducted by the Institute of Education, University of London completed its work in summer 2004. The evaluation report (Hallam et al., 2004) reported the following findings for the impact on pupils:

- a substantial increase in perceptions of pupils' engagement with learning, with particular notable impact on lower attainers, and shy and disengaged pupils in a special school for pupils with complex learning needs;
- better motivation, more positive attitudes to learning, and, for many, enhanced confidence;
- some improvements in behaviour and more cooperation in class in teamwork and in learning;
- dramatic improvements in pupils' learning skills, in learning about their strengths and weaknesses and about what they needed to do to make progress, so encouraging them to take more responsibility for their learning.

For the teachers, they reported greater awareness of the needs of individual pupils, and improvement in their motivation, confidence and enjoyment of their work. They believed that their capacity for self-evaluation, reflection and continuous improvement had been enhanced. A positive impact on their schools as a whole was also reported, and a similar benefit for parents was reported by the primary schools.

Just as these features reflected the experience of the KMOFAP project (which was not independently evaluated), so did most of the points which were judged to contribute to its success. These included the provision of time out of class for teachers to plan, prepare, reflect and evaluate, the action research elements of the project, and the commitment of each school's head teacher and senior management team.

The evaluation also revealed several challenges. One was that some staff found that the initiative called for a fundamental change in their pedagogy, which they found stressful, and for more priority in developing differentiation in implementation of the strategies. The need to meet demands of external accountability was also a cause for

concern, with teachers reporting tension between the demands of summative assessment and the implementation of new formative practices. Again, in all of these features there was close correspondence with the KMOFAP experience.

One problem with the dissemination is that the phrase 'Assessment for Learning' has been attached as a headline label to programmes and publications which do not justify this title, or at the very least, have an idiosyncratic interpretation of the phrase. The most frequent misunderstanding has been to equate this work with frequent summative testing, even to the extent of publicizing highly atomized tests so that pupils' progress towards myriad component targets can be subject to regular checks. A less evident misinterpretation is based on a belief that a short training day, followed up by bulky documentation setting out in detail what teachers should do, will produce the required changes. It will not – the changes needed run too deep for such an approach to work.

## Future issues

Many questions arise from this work that await further research inquiry. Some will be taken further in subsequent chapters of this book. The need to co-ordinate all of the above issues in a comprehensive theoretical framework linking assessment in classrooms to issues of pedagogy and curriculum will be tackled in Chapter 13. The tensions and possible synergies between teachers' own assessments and the assessment results and methods required by society will be explored further in Chapter 15.

A further issue is that of the assumptions about learning underlying the curriculum and pedagogy. The beliefs of teachers about learning, about their roles as assessors and about the 'abilities' and prospects of their students, will affect their interpretations of their students' learning work, and will thereby determine the quality of their formative assessment. This will be taken further in Chapters 12 and 13. A parallel inquiry is also needed into the perceptions and beliefs held by students about themselves as learners, and into their experience of the changes that follow from innovations in formative assessment: exploration of this issue is a current aim of the ESRC-funded project: Learning How to Learn: in Classrooms, Schools and Networks.

Light will also be cast by that project on the problem of the generalizability of the findings from KMOFAP and from the Scottish initiative. The experience so far of schools basing their own innovations on the existing findings of results from research and from recently

developed practice is that a sustained commitment over at least two years is needed, that evaluation and feedback have to be built in to any plan, and that any teachers involved need strong support, both from colleagues and from their school leadership. The more recent Learning to Learn project, a collaboration between Cambridge, King's College and the Open University, has implemented interventions based in part on the findings of our project: this work is reported in Chapter 3, and issues of the leadership of such changes feature in Chapter 4.

Other issues that might repay further exploration are:

- the surprising feature that the research in this field has paid virtually no attention to issues relating to race, class and gender;
- the effect on practice of the content knowledge, and the pedagogical content knowledge, that teachers deploy in particular school subjects: issues for enquiry would be the way in which these resources underlie each teacher's composition and presentation of the learning work, and the interpretative frameworks that he or she uses in responding to the evidence provided by feedback from students;
- the need to pursue in more detail the many issues about pedagogy that are entailed in formative assessment work, notably the deployment in this context of the results of the numerous studies of classroom dialogue (see e.g. Alexander, 2008), and the findings of research and development work aimed at improving the quality of peer-group work (Blatchford et al., 2006; Mercer et al., 2004);
- the nature of the social setting in the classroom, as influenced both by the divisions of responsibility between learners and teachers in formative assessment, and by the constraints of the wider school system.
- the need to extend work of this nature to other groups, notably pupils in infant and junior school and students in post-16, tertiary and non-statutory assessment settings (Chapter 9).

More generally, this work raises questions about the 'application' of research to practice, and the links between this and the professional development of teachers (Black and Wiliam, 2003). Researching how teachers take on research, adapt it, and make it their own is much more difficult than researching the effects of different curricula, of class sizes, or of the contribution of classroom assistants. Furthermore, the criteria applied in judging the practical value of research aligned to development can easily be made too stringent. If, as we believe is reasonable, an approach in which 'the balance of probabilities' rather than 'beyond reasonable doubt' was adopted as the burden of proof,

then this type of educational research would be accepted as having much to say. Thus we take issue with the stance of some policy makers who appear to want large-scale research conducted to the highest standards of analytic rationality, but also insist that the associated findings are also relevant to policy. It may often be the case that these two goals are, in fact, incompatible. To put it another way, when policy without evidence meets development with some evidence, development should prevail.

This chapter is based on a story. We claim that it is an important story, in that the success of the project that it describes helped to give impetus to the wider adoption of formative assessment practices and to recognition of their potential. The significance for this book is that the practices developed with the teachers helped to put classroom flesh on the conceptual bones of the idea of assessment for learning. Given that serving learning is the first and most important of the purposes of assessment, this is an appropriate starting point for the comprehensive picture of assessment that is developed through the subsequent chapters.

## Note

- 1 In addition to *Inside the Black Box* and *Working Inside the Black Box*, there are a further eight booklets exploring aspects of assessment for learning in primary schools, and in specific subjects in secondary schools (currently Design and Technology, English, Geography, ICT, Mathematics, Modern Foreign Languages, and Science). Further details can be found at <http://shop.gl-assessment.co.uk/home.php?cat=383>.

## References

- Alexander, R. (2008) *Towards Dialogic Teaching: Rethinking Classroom Talk* (4th edn). York, UK: Dialogos.
- Assessment Reform Group (ARG) (1999) *Assessment for Learning: Beyond the Black Box*. Cambridge, UK: University of Cambridge School of Education.
- Assessment Reform Group (ARG) (2002a) *Assessment for Learning: 10 Principles*. Cambridge, UK: University of Cambridge Faculty of Education.
- Assessment Reform Group (ARG) (2002b) *Testing, Motivation and Learning*. Cambridge, UK: University of Cambridge Faculty of Education.
- Black, H. (1986) 'Assessment for learning', in D.L. Nuttall (ed.), *Assessing Educational Achievement*. London: Falmer Press. pp. 7–18.
- Black, P. J. (1993) 'Formative and summative assessment by teachers', *Studies in Science Education*, 21(1): 49–97.
- Black, P. and Wiliam, D. (1998a) 'Assessment and classroom learning', *Assessment in Education: Principles, Policy and Practice*, 5(1): 7–73.

- Black, P. and Wiliam, D. (1998b) *Inside the Black Box: Raising Standards through Classroom Assessment*. London, UK: King's College London School of Education.
- Black, P. and Wiliam, D. (2003) 'In praise of educational research: formative assessment', *British Educational Research Journal*, 29(5): 623–37.
- Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2002) *Working Inside the Black Box: Assessment for Learning in the Classroom*. London, UK: GL Assessment.
- Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2003). *Assessment for Learning: Putting it into Practice*. Buckingham, UK: Open University Press.
- Black, P., Harrison, C., Hodgen, J., Marshall, M. and Serret, N. (2010) 'Validity in teachers' summative assessments', *Assessment in Education* 17(2): 215–32.
- Blatchford, P., Baines, E., Bassett, P., Chowne, A. and Rubie-Davies, C. (2006). 'The effect of a new approach to group work on pupil–pupil and teacher–pupil interactions', *Journal of Educational Psychology*, 98(4): 750–65.
- Brousseau, G. (1984). 'The crucial role of the didactical contract in the analysis and construction of situations in teaching and learning mathematics' (G. Seib, trans.), in H.-G. Steiner (ed.), *Theory of Mathematics Education: ICME 5 topic area and miniconference* (Vol. 54). Bielefeld, Germany: Institut für Didaktik der Mathematik der Universität Bielefeld. pp. 110–19.
- Butler, R. (1988). 'Enhancing and undermining intrinsic motivation: the effects of task-involving and ego-involving evaluation on interest and performance', *British Journal of Educational Psychology*, 58: 1–14.
- Chadwick, E. B. (1864). 'Statistics of educational results', *Museum: a Quarterly Magazine of Education, Literature, and Science*, 3: 479–84.
- Crooks, T. J. (1988). 'The impact of classroom evaluation practices on students', *Review of Educational Research*, 58(4): 438–81.
- Dweck, C. S. (2000). *Self-theories: Their Role in Motivation, Personality and Development*. Philadelphia, PA: Psychology Press.
- Flyvbjerg, B. (2001). *Making Social Science Matter: Why Social Inquiry Fails and How it can Succeed Again*. Cambridge, UK: Cambridge University Press.
- Foos, P.W., Mora, J.J. and Tkacz, S. (1994) 'Student study techniques and the generation effect', *Journal of Educational Psychology*, 86(4): 567–76.
- Ginsburg, H.P. (2001) *The Mellon Literacy Project: What Does it Teach Us about Educational Research, Practice, and Sustainability?* New York, NY: Russell Sage Foundation.
- Hallam, S., Kirton, A., Peffers, J., Robertson, P. and Stobart, G. (2004) *Evaluation of project 1 of the Assessment is for Learning Development Programme: Support for Professional Practice in Formative Assessment. Final report*. London, UK: Institute of Education, University of London.
- Hargreaves, D.H. (1999) 'The knowledge-creating school', *British Journal of Educational Studies*, 47(2): 122–44.
- King, A. (1992) 'Facilitating elaborative learning through guided student-generated questioning', *Educational Psychologist*, 27(1): 111–26.
- Kluger, A.N. and DeNisi, A. (1996) 'The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory', *Psychological Bulletin*, 119(2): 254–84.
- Lee, C. and Wiliam, D. (2005) 'Studying changes in the practice of two teachers developing assessment for learning', *Teacher Development: an International Journal of Teachers' Professional Development*, 9(2): 265–83.
- Linn, R.L. (1989) 'Current perspectives and future directions', in R.L. Linn (ed.), *Educational Measurement* (3rd edn). Washington, DC: American Council on Education/Macmillan. pp. 1–10.

- Mansell, W., James, M. and Assessment Reform Group (2009) *Assessment in Schools: Fit for Purpose?* London, UK: Economic and Social Research Council Teaching and Learning Research Programme.
- Mercer, N., Dawes, L., Wegerif, R. and Sams, C. (2004) 'Reasoning as a scientist: ways of helping children to use language to learn science', *British Educational Research Journal*, 30(3): 359–77.
- National Curriculum Task Group on Assessment and Testing (1988) *A Report*. London, UK: Department of Education and Science.
- Natriello, G. (1987) 'The impact of evaluation processes on students', *Educational Psychologist*, 22(2): 155–75.
- Perrenoud, P. (1991) 'Towards a pragmatic approach to formative evaluation', in P. Weston (ed.), *Assessment of Pupil Achievement* (Vol. Part A: 25). Amsterdam, Netherlands: Swets & Zeitlinger. pp. 79–101.
- Ramaprasad, A. (1983) 'On the definition of feedback', *Behavioural Science*, 28(1): 4–13.
- Reeves, J., McCall, J. and MacGilchrist, B. (2001) 'Change leadership: planning, conceptualization and perception', in J. MacBeath and P. Mortimore (eds), *Improving School Effectiveness*. Buckingham, UK: Open University Press. pp. 122–37.
- Rowe, M.B. (1974) 'Wait time and rewards as instructional variables, their influence on language, logic and fate control', *Journal of Research in Science Teaching*, 11: 81–94.
- Sadler, D.R. (1989) 'Formative assessment and the design of instructional systems', *Instructional Science*, 18: 119–44.
- Sadler, D.R. (1998) 'Formative assessment: revisiting the territory', *Assessment in Education: Principles, Policy and Practice*, 5(1): 77–84.
- Toulmin, S. (2001) *Return to Reason*. Cambridge, MA: Harvard University Press.
- Vispoel, W.P. and Austin, J.R. (1995) 'Success and failure in junior high school: a critical incident approach to understanding students' attributional beliefs', *American Educational Research Journal*, 32(2): 377–412.
- White, B.Y. and Frederiksen, J.R. (1998) 'Inquiry, modeling, and metacognition: making science accessible to all students', *Cognition and Instruction*, 16(1): 3–118.
- Wiliam, D. (2003) 'The impact of educational research on mathematics education', in A. Bishop, M.A. Clements, C. Keitel, J. Kilpatrick and F.K.S. Leung (eds), *Second International Handbook of Mathematics Education*. Dordrecht, Netherlands: Kluwer Academic Publishers. pp. 468–88.
- Wiliam, D., Lee, C., Harrison, C. and Black, P. J. (2004) 'Teachers developing assessment for learning: impact on student achievement', *Assessment in Education: Principle, Policy and Practice*, 11(1): 49–65.
- Wilson, S.M. and Berne, J. (1999) 'Teacher learning and the acquisition of professional knowledge: an examination of research on contemporary professional development', in A. Iran-Nejad and P. D. Pearson (eds), *Review of Research in Education*, 24. Washington, DC: American Educational Research Association. pp. 173–209.