OECD CERI

WHAT WORKS IN INNOVATION IN EDUCATION

CHANGING TEACHING THROUGH FORMATIVE ASSESSMENT: RESEARCH AND PRACTICE THE KING'S-MEDWAY-OXFORDSHIRE FORMATIVE ASSESSMENT PROJECT

By

Paul Black and Dylan Wiliam, King's College, London[1]

NOT FOR CITATION

[1] We would like to acknowledge the initiative of the Assessment Policy Task Group of the British Educational Research Association (now known as the Assessment Reform Group) who gave the initial impetus and support for our research review. We are grateful to the Nuffield Foundation who funded the original review, and the first phase of our project. We are also grateful to Professor Myron Atkin and his colleagues in Standford University, who secured funding from the US National Science Foundation (NSF) for the last phase. Finally, we are indebted to the Medway and Oxfordshire local education authorities, their six schools and above all to their 36 teachers who took on the central and risky task of turning our ideas into practical working knowledge.

Introduction

This paper is the story of a development which started with a review of what research had to say about formative assessment. The work of this review is 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 in a second section is followed by reflections on outcomes and implications in a third section. Broader reflections on how this experience throws light on the task of turning research results into practice are set out in a fourth section.

The research review

The story starts with our long-standing interest in formative assessment, which led us to decide that it was essential to review the literature in order to look for evidence that improving formative assessment raises standards. It also seemed necessary to look both for evidence about whether or not present practice left room for improvement, and for guidance about how to improve formative assessment.

Our survey of the research literature involved checking through many books, through the issues of over 160 journals for a period of nine years, and studying earlier reviews of research (Crooks, 1988; Natriello, 1987). This process yielded about 580 articles or chapters to study. Out of this we have prepared a lengthy review, which used material from 250 of these sources. The review was published (Black and Wiliam, 1998a) together with comments on our work by experts from five different countries.

A ***first*** section of the review surveyed the evidence. An example was a study published in 1986, which concentrated – but not exclusively – on classroom assessment work for children with mild handicaps, and surveyed a large number of formative innovations from which 23 were selected (Fuchs and Fuchs, 1986). All in this group showed quantitative evidence of learning gains by comparing data for an experimental group with similar data from a control group. Since then, many more papers have been published describing similarly rigorous quantitative experiments. Our own review reported about 20 more such studies all of which showed that innovations which include strengthening the practice of formative assessment produced significant, and often substantial, learning gains. These studies 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 which had, as a common feature, enhanced formative assessment indicated that it is this feature

which 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.

A ***second*** section covered research into current practices of teachers. The picture that emerged was depressing. 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 emphasising 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. Overall it seemed that formative assessment was weak in practice and that its implementation calls for rather deep changes both in teachers' perceptions of their own role in relation to their students and in their classroom practice.

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 to take action, their selection of a line of action and the nature of their commitment to it. 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 (ie the 'mastery' level), that such feedback was given rapidly; 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. A notable example was the extensive review of empirical evidence by Kluger and DeNisi (1996) which showed that feedback can have positive effects only if the feedback is formulated and used as a guide to improvement. Of equal importance was the conceptual analysis which defined feedback as "… information about the gap between the actual level and the reference level of a system parameter which is used

to alter the gap in some way" (Ramaprasad, 1983) and the development of this by Sadler (1989) to emphasise that learners must understand both the "reference level" – *i.e.* the goal of their learning – and the actual level of their understanding.

Equally important was the clear message from the research on attribution theory (for example by Vispoel and Austin, 1995) 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).

Overall, the features which seem to characterise many of the studies were:

- 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.
- Underlying the various approaches are assumptions about what makes for effective learning – in particular that students have to be actively involved.
- 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.
- 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.

Interpreting the research

Synthesising research cannot be an objective process – it will inevitably remain subjective. The structure of the six sections outlined above did not emerge automatically: it was our chosen way to reconceptualise, to organise, and to focus the relevant literature field. Our definition of "relevance" expanded as we went along, so we had to find ways of organising a widening field of research, and to make new conceptual links in order to be able to combine the various findings into as coherent a picture as possible. This was one reason why our review generated a momentum for work in this field: it provided a new framework that would be difficult to create in any other way. Reviewing research is not merely a derivative form of scholarship.

Publicity

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 publicised 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 characterised as those of "academic rationality", the standard for *Inside the Black Box* was much closer to that of "reasonableness" advocated by Stephen Toulmin for social enquiry (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* has been as much due to its rhetorical force as to its basis in evidence. 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).

Moving into action

Setting up a project

The second stage of our story followed the first almost inevitably: given that our review had shown that innovations in formative assessment could raise standards of student achievement, it was natural to think about ways to help schools secure these benefits. Our own experience of teachers' professional development had taught us that the implementation of new practices in classrooms could not be a straightforward matter of setting out a recipe for teachers to follow. For one reason, given the varied nature of the innovations and the different contexts in which they had been tried, we could not assume that they could simply be "copied" to other contexts. A second reason was that, from reading the reports of the researchers, one could not describe their work at the level of detail that would be needed to formulate advice on how to replicate them. A third reason, which would have been decisive even in the absence of the first two, was our approach to the task of turning research into practice. We believed that new ideas about teaching and learning can only be made to work in particular contexts, in our case that of teachers in (initially) UK secondary schools, if teachers are able to transform them and so create new practical knowledge relevant to their task.

So we obtained funding (from the UK's Nuffield Foundation) for a two-year development project. Six schools who taught students in the age range 11 to 18 years agreed to collaborate with us: each selected two science and two mathematics teachers willing to take on the risks and extra work involved. In second year of the project we added two teachers of English, from each of same schools, and one additional mathematics and science teacher, so that in all 48 teachers were involved. They were supported by staff from their local (district) education authorities and the project was called the King's-Medway-Oxfordshire Formative Assessment Project (KMOFAP) to highlight our close collaboration with all the other partners (Black and Wiliam, 2003).

The teachers and the researchers met in a whole day meeting every five weeks, over 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 both 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, and on a few discussions with student groups.

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 200 such contributions. These have ranged across all subjects, and across both primary and secondary 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 United Kingdom, with education ministries in Scotland and in Jersey, and in a recent exploration of classroom outcomes for a government programme which aims to improve teaching and learning practices in schools.

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 – the achievement data used were 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.

The practices developed

These practices will be described here under four headings: oral feedback in *classroom 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).

For *classroom dialogue* the aim was to improve the interactive feedback which is central to formative assessment. An account of wait time research (Rowe, 1974) motivated teachers to allow 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. One particular way to increase participation was to ask students to brainstorm ideas, perhaps in pairs, for two to three minutes prior to the teacher asking for contributions. Then all answers, right or wrong, had to 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 mis-conceptions in that knowledge, so that their next moves could address the learners' real needs.

As they tried to develop this approach, teachers realised that more effort had to be spent in framing questions that were worth asking, *i.e.* questions which explored issues that are critical to the development of students' understanding. They also had to focus closely on follow-up activities to formulate meaningful responses and challenges that would help students to extend their 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, not least because it felt at first as if they were losing control. 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 which have established that, whilst 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 experience showed that the provision of comments gave both students and their parent's advice on how to improve. It also set up a new focus on the learning issues 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, for it was now evident that these had to identify what had been done well and what still needed improvement, and to give guidance on how to make that improvement. As the skills of formulating and using such feedback were developed, it became more clear that the quality of the tasks set for written homework or class-work was critical: such tasks, alongside oral questioning, had to be designed to encourage students to develop and express their understanding of the key features of what they had learnt.

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. Insofar 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).

For the development of self-assessment skills, the first and most difficult task is to get students to think of their work in terms of a set of goals. In practice, peer-assessment turned out to be an important stimulus to self-assessment. Peer-assessment is uniquely valuable because students may accept, from one another, criticisms of their work which they would not take seriously if made by their teacher. Peer work is also valuable because the interchange will be in language that students themselves would naturally use, and because students learn by taking the roles of teachers and examiners of others (Sadler, 1998). In particular, students appear to find it easier to make sense of criteria for their work in the context of other students' work than when looking at their own.

However, for such peer-group work to succeed, many students needed guidance about how to behave in groups, *e.g.* in listening to one another, taking turns, and offering affirmation together with constructive criticism about one another's work. A typical exercise would be on the marking of homework. Students were asked to label their work with "traffic lights", *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 realised that many students had no strategy for preparing for a test by 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 give particular difficulty.

A further idea was introduced from research studies (Foos *et al*., 1994; King, 1992) which 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. 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 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 test set and marked externally.

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 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. We had, after all, stressed that feedback functioned formatively only if the information fed back to the learner was used by the learner in improving performance. But whilst one can work out after the event whether or not any feedback has had the desired effect, what the teachers 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.

So 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 which 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 must be active in creating their own understandings.

These ideas reflect some of the main principles of the constructivist view of learning – to start where the students are and to involve the students actively in the process. It became clear to the teachers that, no matter what the pressure to achieve good test and examination scores, learning cannot be done for the student; it has to be done by the student.

Students came 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 others 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, all within a change in the classroom culture to which all four activities contributed, were creating more a more rich community of learners where the social learning of students would become more salient and effective.

A learning environment and changes of role

There are also deeper issues here. A learning environment has to be "engineered" to involve students more actively in the tasks. The emphasis has to be on the students doing 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

This teacher had changed his role, from presenter of content to 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. Toward the end of the project, they described this same process not as a loss of control, but 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) put it:

> Every teacher who wants to practice 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

emphasising 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?”

What has been happening here is that everybody’s role expectations, *i.e.* what teachers and students think that being a teacher or being a student requires you to do, have 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 and steadily, as experience developed and confidence grew in the use of the various strategies for enriching feedback and interaction.

Further research

In our 1998 review, we listed a number of issues for study by further research. The first issue was the extent to which the context of any study is artificial so that generalisability of the results cannot be guaranteed. This reservation was one of the reasons why we developed the KMOFAP work and now it can be applied to the generalisability of the findings of that study. Our experience of seeing other schools base their own innovations on the KMOFAP results is that a sustained commitment over at least two years is needed, that evaluation and feedback have to be built into any plan, and that any teachers involved need strong support, both from colleagues and from their school leadership.

A second research interest arose from a surprising feature – that the research we studied seemed to pay no attention to issues relating to race, class and gender; these issues still await exploration. A third area for further enquiry is that of beliefs and assumptions about learning theory. Both 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. A parallel enquiry is 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.

A fourth area is the effect on practice of the content knowledge, and the pedagogical content knowledge, that teachers deploy in their 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 social setting of a classroom, the community it forms, and the quality of the interactions within that community, all have strong effects in such innovations as

better classroom dialogue and peer- and self-assessment. Matters to be studied here would 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.

Two further issues now seem important. One is the tensions and possible synergies between teachers' own assessments and the assessment results and methods required by society. The other is 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 – a task which remains to be tackled.

Research and practice

Why did it work?

At one level, our story was now complete. A basis in research had led to a successful innovation and the publication of its outcomes proved as popular as the original report of the research (Black *et al.*, 2002, 2003). However, we were surprised that it had been so successful in promoting quite radical changes in teachers' practice, and wondered whether lessons could be learnt from it about the notoriously difficult problem of turning research into practice.

One 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: where educational research is concerned, the facts do not necessarily speak for themselves. 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.

A further relevant factor about the content is that the ideas had an intrinsic acceptability to the teachers. We were talking about improving learning in the classroom, which was central to their professional identities, as opposed to bureaucratic measures such as target-setting. 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*.

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 failed attempts to change the *system*, in the 80s and 90s in England, were set aside. Whilst it might have been merely prudent to not try again to tilt at windmills, the more fundamental strength was that it was at the level chosen, that of the core of learning, 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. The evidence of learning gains, from the literature review and from our project, restates and reinforces the claim for priority of formative work that earlier policy recommendations (DES, 1988) tried in vain to establish. The debate about how policy should secure optimum synergy between teachers' formative, teachers' summative, and external assessments is still unresolved, but the new salience of work on formative assessment has now shifted the balance of the arguments.

The process strategy

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 acknowledgement that teachers need time, freedom, and support from colleagues, in order to reflect critically upon and to develop their practice (Lee, 2005), 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 realised, traditional research designs – in which teachers are "told" what to do by researchers – would not be appropriate. 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 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 re-negotiate 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.

Through our work with teachers, we have come to understand more clearly how the task of applying research into 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 were leaving them to work it out for themselves. As they came to know us better, they realised that, at the level of everyday classroom practice, we really did not know what to do.

Making research practical

Whilst we do not believe that all educational research should be useful, we do believe strongly that the majority of research in education should be undertaken with a view to improving educational provision – research in what Stokes (1997) calls "Pasteur's quadrant". And although we do not yet know everything about "what works" in teaching, we believe that there is a substantial consensus on the kinds of classrooms that promote the best learning. What we know much less about is how to get this to happen.

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. While we do not know as much as we would like to know about effective professional development, if we adopt "the balance of probabilities" rather than "beyond reasonable doubt" as our burden of proof, then educational research has much to say. When policy without evidence meets development with some evidence, development should prevail. 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 the findings of which are also relevant to policy. It may often be the case that these two goals are, in fact, incompatible.

REFERENCES

Black, P. and D. Wiliam (1998a), "Assessment and Classroom Learning", *Assessment in Education,* Vol. 5, pp. 7-71.

Black, P. and D. Wiliam (1998b), *Inside the Black Box: Raising Standards through Classroom Assessment*. See also *Phi Delta Kappan,* Vol. 80, pp. 139-148.

Black, P. and D. Wiliam (2003), "In Praise of Educational Research: Formative Assessment", *British Educational Research Journal*, Vol. 29(5), pp. 623-637.

Black, P. *et al.* (2002), *Working inside the Black Box: Assessment for Learning in the Classroom*, Department of Education and Professional Studies, King's College, London.

Black, P. *et al.* (2003), *Assessment for Learning: Putting it into Practice*, Open University Press, Buckingham, United Kingdom.

Brousseau, G. (1984), "The Crucial Role of the Didactical Contract in the Analysis and Construction of Situations in Teaching and Learning Mathematics" in H.-G. Steiner (ed.), *Theory of Mathematics Education: ICME 5 Topic Area and Miniconference*, Institut für Didaktik der Mathematik der Universität Bielefeld, Bielefeld, Germany, pp. 110-119.

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*, vol. 58, pp. 1-14.

Crooks, T.J. (1988), "The Impact of Classroom Evaluation Practices on Students", *Review of Educational Research,* Vol. 58, pp. 438-481.

DES (1988), *Task Group on Assessment and Testing: A Report*, Department of Education and Science and the Welsh Office, London.

Flyvbjerg, B. (2001), *Making Social Science Matter: Why Social Inquiry Fails and how it can Succeed again*, Cambridge University Press, Cambridge UK.

Foos, P.W., J.J. Mora and S. Tkacz (1994), "Student Study Techniques and the Generation Effect", *Journal of Educational Psychology*, Vol. 86, pp. 567-576.

Fuchs, L.S. and D. Fuchs (1986), "Effects of Systematic Formative Evaluation: a Meta-Analysis", *Exceptional Children*, Vol. 53, pp. 199-208.

Hargreaves, D.H. (1999), "The Knowledge Creating School", *British Journal of Educational Studies*, Vol. 47, pp. 122-144.

King, A. (1992), "Facilitating Elaborative Learning through Guided Student-generated Questioning", *Educational Psychologist*, Vol. 27, pp. 111-126.

Kluger, A.N. and A. DeNisi (1996), "The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory", *Psychological Bulletin*, Vol. 119, pp. 254-284.

Lee, C. (2005), "Studying Changes in the Practice of Two Teachers", *International Journal of Teacher Development*, in press.

Natriello, G. (1987), "The Impact of Evaluation Processes on Students", *Educational Psychologist*, Vol. 22, pp. 155-175.

Perrenoud, P. (1991), "Towards a Pragmatic Approach to Formative Evaluation", in P. Weston (ed.), *Assessment of Pupils Achievement: Motivation and School Success*, Swets and Zeitlinger, Amsterdam, pp. 79-101.

Ramaprasad, A. (1983), "On the Definition of Feedback", *Behavioral Science*, Vol. 28, pp. 4-13.

Reeves, J., J. McCall and B. MacGilchrist (2001), "Change Leadership: Planning, Conceptualization and Perception", in J. MacBeath and P. Mortimore (eds.), *Improving school effectiveness*, Open University Press, Buckingham UK, pp. 122-137.

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*, Vol. 11, pp. 81-94.

Sadler, R. (1989), "Formative Assessment and the Design of Instructional Systems", *Instructional Science*, Vol. 18, pp. 119-144.

Sadler, R. (1998), "Formative Assessment: Revisiting the Territory", *Assessment in Education*, Vol. 5, pp. 77-84.

Stokes, D.E. (1997), *Pasteur's Quadrant: Basic Science and Technological Innovation*, Brookings Institution Press, Washington DC.

Toulmin, S. (2001), *Return to Reason*, Harvard University Press, Cambridge MA.

Vispoel, W.P. and J.R. Austin (1995), "Success and Failure in Junior High School: A Critical Incident Approach to Understanding Students' Attributional Beliefs", *American Educational Research Journal*, Vol. 33, pp. 377-412.

White, B.Y. and J.R. Frederiksen (1998), "Inquiry, Modeling, and Metacognition: Making Science Accessible to all Students", *Cognition and Instruction*, Vol. 16, pp. 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*, Kluwer Academic Publishers, Dordrecht Netherlands, pp. 469-488.

Wiliam, D. *et al*. (2004), "Teachers Developing Assessment for Learning: Impact on Student Achievement", *Assessment in Education*, Vol. 11, pp.49-65.

Wilson, S.M. and J. Berne (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*, American Educational Research Association, Washington DC, pp. 173-209