DRAFT

The Nature and Value of Formative Assessment for Learning

Paul Black

with the King's College London Assessment for Learning Group
(Christine Harrison, Clare Lee[1], Bethan Marshall, Dylan Wiliam[2])

Abstract

The this paper has two foci. The first is to present an account of how we developed formative assessment practices with a group of 36 teachers. This is then complemented by a reflection on the productive and positive experience of these teachers, in the light of learning principles, of changes in the roles of teachers and pupils in the task of learning, and of effects on the self-esteem and motivation of pupils. Attention then shifts to the second focus, which is on the ways in which these teachers struggled with the interface between formative assessment and summative testing. The conclusion is that the potential of enhanced classroom assessment to raise standards may never be fully realised unless the regimes of assessment for the purposes of accountability and certification of pupils are reformed.

The starting point

In 1998 we published a review summarising the results from over 250 articles by researchers from several countries. Our main purpose was to survey the evidence about the effects of improving formative assessment on pupils' performance (Black & Wiliam, 1998a). This review showed that there was a strong body of evidence to support a claim that formative assessment practices can raise standards.

At the same time, the published evidence also showed that such practices were only weakly developed in most classrooms. Three main problems were identified. The first was that the assessment methods that teachers use are not effective in promoting good learning.The second was that marking and grading practices tend to emphasise competition rather than personal improvement. The third problem was that assessment feedback often has a negative impact, particularly on pupils with low attainments who are led to believe that they lack 'ability' and are not able to learn.

However, whilst these conclusions were a spur to action, they could not provide recipes for improvement. The reported surveys and experiments lacked the detail that would enable teachers to implement the practices in classrooms. It seemed to us then that teachers would need a variety of living examples of implementation, by teachers with whom they could identify and from whom they could both derive conviction and confidence that they could do better.

[1] Now assessment adviser with the Warwickshire LEA

[2] Now with the Educational Testing Service, Princeton, NJ, USA

To publicise the results we also published a 20-page booklet *Inside the Black Box: Raising standards through classroom assessment.* (Black & Wiliam, 1998b). Over 40000 copies have been sold and the work is widely quoted. Thus it was clear that teachers were interested in what we had to say.

So we decided to start on the exploratory development work that was called for. To do this, we needed to collaborate with a group of teachers willing to take on the risks and extra work involved, and to secure support from their schools and their Local Education (district) Authorities (known in the UK as LEAs). The funding for the project was provided through the generosity of the Nuffield Foundation. We were fortunate to find, in the Medway and Oxfordshire LEAs, advisory staff who understood the issues and who were willing to work with us. Each authority selected three secondary schools, spanning a range of catchment backgrounds; they included one boys' and one girls' school, the other four being mixed. Each school selected two science and two mathematics teachers. We discussed the plans with the head of each school, and then called the first meeting of the 24 teachers in January 1999 . In the subsequent second year of the project we added 12 teachers of English, two from each of same schools. The project became known as the King's -Medway -Oxfordshire Formative Assessment Project (KMOFAP). The findings from this work have been published both as a short booklet for teachers (Black et al. 2002) and as a book for teachers (Black et al. 2003)

The ways in which the partners involved worked together has been written up elsewhere (Black & Wiliam 2003). For the present purpose, it is the outcomes that are important. The findings presented here are based on the observations and records of visits to classrooms by the King's team, records of meetings of the whole group of teachers, interviews with and writing by the teachers themselves, and a few discussions with pupil groups.

Throughout the development of the project, members of the King's team have responded to numerous invitations to talk to other groups of teachers and advisers: over three years they have made over 100 such contributions. These have ranged across all subjects, and across both primary and secondary phases. In addition, there has been sustained work with some primary schools. All of this makes us confident that our general findings will be of value to all, although some important details may differ between different age groups and different subjects. Some of the results of our work have been incorporated into the Key Stage 3 initiative and into the primary strategy, whilst researchers and teachers in the project have been appointed consultants to national programmes in Scotland and in Jersey to develop assessment for learning.

The concept

In our short publication for teachers (Black et al. 2002) we included the following extended definition of our idea of formative assessment :

> *Assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting pupils' learning. It thus differs from assessment designed primarily to serve the purposes of accountability, or of ranking, or of certifying competence.*

> *An assessment activity can help learning if it provides information to be used as feedback, by teachers, and by their pupils in assessing themselves and each other, to modify the teaching and learning activities in which they are engaged. Such assessment becomes 'formative assessment' when the evidence is actually used to adapt the teaching work to meet learning needs.*

The detail and precision here have been found essential, because many teachers and researchers seem to have misunderstood the term. One example of misunderstanding is that it is about any assessment conducted by teachers, and in particular that giving a test every week and telling the pupils the marks constitutes formative assessment. It does not. Unless some learning action follows from the outcomes, such practice is merely frequent summative assessment. Another example is the belief that it includes portfolio assessment where that is developed with the aim of replacing or supplementing the information which externally imposed tests are designed to provide. Again, there is no formative assessment in such practice except insofar as there is active feedback to change and improve pupils' work as the portfolios are built up. In general, any test or assessment at the end of a piece of learning is too late for formative purposes, precisely because it is at the end, so there is no opportunity to use its results for feedback to improve performance of the pupils involved.

The learning gains

Although our review of the international research literature showed that enhanced formative assessment produces gains in pupil achievement, we were clear that it was important to have some indication of the kinds of gains that could be achieved in normal classrooms, and over an extended period of time. The criteria for 'normal' were that the content of the teachers' lessons would be that of their normal plans, and that the only measures of pupils' performance would be those derived from their normal testing procedures. Since each teacher in the project was free to decide the class with which they would work on these ideas, we discussed with each teacher individually how to set up a 'mini-experiment' for each teacher using data available within the school.

Each teacher decided what was to be the 'output' measure for their class, using, as appropriate, data from national examinations and tests and/or results of the school's own tests. For each project class, the teacher identified a control class. In some cases this was a parallel class taught by the same teacher in previous years (and in one case in the same year). In other cases, we used a parallel class taught by a different teacher and, failing that, a non-parallel class taught by the same or different teacher. Where the project and the control classes were not strictly parallel, we controlled for possible differences in ability by the use of 'input' measures, such as scores on a standard Cognitive Abilities Test, or school test scores from the previous year. The data covered changes over one whole school year (for a full report see Wiliam et al. 2004).

To enable us to aggregate the results across the teachers, we used the standardised effect size. For the nineteen teachers on whom we had reliable data, the average effect size was around 0.3. Such improvement, produced across a school, would raise a school in the lower quartile of the UK national performance tables to well above average. It was clear, therefore, that, far from having to choose between teaching well and getting good test results, teachers can actually improve their pupils' results by working with the ideas that developed in our

project.

The findings: how change can happen

This section sets out the main findings about classroom work under four headings: questioning, feedback through marking, peer- and self-assessment, and the formative use of summative tests. Most of the quotations are taken from pieces written by the teachers: the names of the teachers and of the schools are pseudonyms.

Questioning

Many teachers do not plan and conduct classroom dialogue in ways that might help pupils to learn. Research has shown that many leave less than one second after asking a question before, if no answer is forthcoming, asking another question, or answering their own question (Rowe, 1974). A consequence of such short 'wait time' is that the only questions which 'work' are those which can be answered quickly without thought, i.e. questions which call for memorised facts. In consequence, the dialogue is at a superficial level. The key to changing such a situation is to allow longer wait time. Many of the teachers found it hard to do this—they had to break their established habits and, as they changed, the expectations of their pupils were challenged, and their persistence in silence sometimes tested the teacher's resolve. One teacher summarised the overall effects of her efforts to improve the use of question and answer dialogue in the classroom as follows:

Questioning

•My whole teaching style has become more interactive. Instead of showing how to find solutions, a question is asked and pupils given time to explore answers together. My Year 8 target class is now well-used to this way of working. I find myself using this method more and more with other groups.

No hands

•Unless specifically asked pupils know not to put their hands up if they know the answer to a question. All pupils are expected to be able to answer at any time even if it is an 'I don't know'.

Supportive climate

•Pupils are comfortable with giving a wrong answer. They know that these can be as useful as correct ones. They are happy for other pupils to help explore their wrong answers further.

Nancy, Riverside School

Increasing the wait time can lead to more pupils being involved in question and answer discussions, and to an increase in the length of their replies. One particular way to increase participation was to ask pupils to brainstorm ideas, perhaps in pairs, for two to three minutes prior to the teacher asking for contributions. Overall, a consequence of such changes was that teachers learnt more about the pre-knowledge of their pupils, and about any gaps and mis-conceptions in that knowledge, so that their next moves could address the learners' real needs. To exploit such changes they found it necessary to move away from the routine of

limited factual questions and to refocus attention on the quality and the different functions of classroom questions. The questions themselves then became a more significant part of teaching, with attention focused on how they could be constructed and used to explore and then develop pupils' learning. Collaboration between teachers to exchange ideas and experiences about good questions became a priority for the agenda of the project meetings.

However, if this is to be productive, both the responses that the task might evoke and ways of following up these responses have to be anticipated Effective questioning is an important aspect of the impromptu interventions teachers make once the pupils are engaged in an activity. These often include simple questions such as "Why do you think that?" or "How might you express that?", or—in the 'devil's advocate' style— "You could argue that...". This type of questioning became part of the interactive dynamic of the classroom and provided an invaluable opportunity to extend pupils' thinking through immediate feedback on their work.

> *When I really stopped to think I realised that I could make a very large difference to the girls' learning by using all their answers to govern the pace and content of the lesson.*
>
> *Gwen, Waterford School*

Overall the main suggestions for action that have emerged from the teachers' experience are:

- More effort has to be spent in framing questions that are worth asking, i.e. questions which explore issues that are critical to the development of pupils' understanding.
- Wait time has to be increased to several seconds in order to give pupils time to think and everyone should be expected to have an answer and to contribute to the discussion. Then all answers, right or wrong, can be used to develop understanding.
- Follow-up activities have to be rich, in that they provide opportunities to ensure that meaningful interventions that extend the pupils' understanding can take place.
- Pupils become more active as participants, and come to realise that learning may depend less on their capacity to spot the right answer and more on their readiness to express and discuss their own understanding

Put simply, the only point of asking questions is to raise issues about which the teacher needs information or about which the pupils need to think.

Feedback through marking

It is the nature, rather the amount, that is critical when giving pupils feedback on both oral and written work. The research by Butler (1988) established that, whilst pupils' learning can be advanced by feedback through comments, the giving of marks—or grades—has a negative effect in that pupils ignore comments when marks are also given. These results surprised our teachers, but those who abandoned the giving of marks found that their experience confirmed the findings: pupils do engage more productively in improving their work.

Many of the teachers were concerned about the effect of returning pupils' work with comments but no marks and there were conflicts with school policy:

> *My marking has developed from comments with targets and grades, which is the school policy, to comments and targets only. Pupils do work on targets and corrections more productively if no grades are given. Clare [King's researcher] observed on several occasions how little time pupils spend reading my comments if there were grades given as well. My routine is now, in my target class, to;*
>
> *i) not give grades only comments*
>
> *ii) comments highlight what has been done well and what needs further work*
>
> *iii) the minimum follow up work expected to be completed next time I mark the books.*
>
> *Nancy, Riverside School*

Initial fears about how pupils and parents might react turned out to be unjustified, indeed, the provision of comments to pupils helped parents to focus on the learning issues rather than on trying to interpret a mark or grade. The effort that many teachers devote to marking homework may be mis-directed. A numerical mark does not tell pupil how to improve their work, so an opportunity to enhance their learning has been lost.

A policy of improving their comments required more work initially, as teachers had to attend to the quality of the comments that they wrote on pupils' work. here again collaboration between teachers to share examples of effective comments proved very helpful, and experience gradually led to more efficient fluency.

We met a variety of ways of accommodating the new emphasis on comments. Some teachers ceased to assign marks at all, some entered marks in record books but did not write them in the pupils' books, whilst others gave marks only after a pupil had responded to their comments. Some teachers spent more time on certain pieces of work to ensure that they gave good feedback and, to make time for this, either did not mark some pieces, or marked only a third of their pupils' books each week, or involved the pupils in checking straightforward tasks.

As they tried to create useful feedback comments, many of the project teachers realised that they needed to reassess the work that they had asked pupils to undertake. They found that some tasks were useful in revealing pupils' understandings and mis-understandings, but that others focused mainly on conveying information. So some activities were eliminated, others modified, and new and better tasks actively sought.

Overall the main ideas for improvement can be summarised as follows:

- Written tasks, alongside oral questioning, should encourage pupils to develop and show understanding of the key features of what they have learnt.

- Comments should identify what has been done well and what still needs improvement,

and give guidance on how to make that improvement.

- Opportunities for pupils to follow up comments should be planned as part of the overall learning process.

The central point here is that, to be effective, feedback should cause thinking to take place.

Implementation of such reforms can change the attitudes of both teachers and pupils to written work: pupils will be see an assessment of their work less as a competitive and summative judgement and more as a distinctive step in the process of learning. The importance of this aspect will be taken later in this paper.

Peer-assessment and self-assessment

Pupils can only achieve a learning goal if they understand that goal and can assess what they need to do to reach it. So self-assessment is essential to learning (Sadler, 1989). Many who have tried to develop self-assessment skills have found that the first and most difficult task is to get pupils to think of their work in terms of a set of goals. Insofar as they do so they begin to develop an overview of that work so that it becomes possible for them to manage and control it for themselves: in other words, they are developing the capacity to work at a meta-cognitive level.

In practice, peer-assessment turned out to be an important complement to self-assessment. Peer-assessment is uniquely valuable because pupils 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 pupils themselves would naturally use, and because pupils learn by taking the roles of teachers and examiners of others (Sadler, 1998):

> *As well as assessing and marking (through discussion and clear guidance) their own work they also assess and mark the work of others. This they do in a very mature and sensible way and this has proved to be a very worthwhile experiment. The students know that homework will be checked by themselves or another girl in the class at the start of the next lesson. This has lead to a well-established routine and only on extremely rare occasions have students failed to complete the work set. They take pride in clear and well presented work that one of their peers may be asked to mark. Any disagreement about the answer is thoroughly and openly discussed until agreement is reached.*
>
> *Alice, Waterford School*

The last sentence of this quotation brings out an important point—when pupils do not understand an explanation, they are likely to interrupt a fellow pupil when they would not interrupt a teacher. In addition to this advantage, peer-assessment is also valuable in placing the work in the hands of the pupils. Teachers found that they could be more free to observe and reflect on what was happening and to frame helpful interventions.

For such peer-group work to succeed, many pupils will need guidance about how to behave in groups, e.g. in listening to one another and taking turns. Both self- and peer-assessment will only happen if teachers help their pupils, particularly the low-attainers, to develop the

skill. This can take time and practice.

Our experience of work on this theme leads to the following recommendations for improving classroom practice:

- The criteria for evaluating any learning achievements must be made transparent to pupils to enable them to have a clear overview both of the aims of their work and of what it means to complete it successfully. Such criteria may well be abstract—concrete examples should be used in modelling exercises to develop understanding.

- Pupils should be taught the habits and skills of collaboration in peer-assessment, for these are of intrinsic value. Futhermore, peer discussion can help self-assessment by helping pupils to see their own work more objectively – through the eyes of their peers.

- Pupils should be encouraged to keep in mind the aims of their work and to assess their own progress to meet these aims as they proceed. They will then be able to guide their own work, and so become independent learners.

The main point here is that peer and self-assessment make unique contributions to the development of pupils' learning—they secure aims that cannot be achieved in any other way.

The formative use of summative tests

The practices of self and peer-assessment were applied to help pupils prepare for examinations, for example in tackling the following problem:

> *They did not mention any of the reviewing strategies we had discussed in class. When questioned more closely it was clear that many spent their time using very passive revision techniques. They would read over their work doing very little in the way of active revision or reviewing of their work. They were not transferring the active learning strategies we were using in class to work they did at home.*
>
> *Tom, Riverside School*

To change this situation, pupils were asked to 'traffic light' a list of key words or the topics on which the test would be set. The 'traffic lights' were icons, used to label their work green, amber, or red according to whether they thought they had good, partial, or little understanding. The point of this was to stimulate the pupils to reflect on where they felt their learning was secure, which they marked green, and where they needed to concentrate their efforts, in amber and red. These traffic lights then formed the basis of a revision plan. Pupils were be asked to identify questions on past examination papers that tested their red areas and then worked with books and in peer groups to ensure they could successfully answer those questions.

The aftermath of tests was also used as an occasion for formative work. Peer marking of test papers was found helpful, as with normal written work, and was particularly useful where pupils were required first to formulate a mark scheme, an exercise which focused attention on criteria of quality relevant to their productions. After peer marking, teachers could

reserve their time for discussion of the questions that had given particular difficulty; peer tutoring was used to tackle those problems encountered by only a minority.

A further idea was been introduced from research studies (Foos et al., 1994; King, 1992) which have shown that pupils 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 called for, and so helped developed, an overview of the topic.

These developments challenge common expectations. Some have argued that formative and summative assessments are so different in their purpose that they have to be kept apart, and such arguments are strengthened by experience of the harmful influence that narrow 'high-stakes' summative tests can have on teaching. Indeed, the King's team advised the project teachers to isolate their formative initiatives from the summative test aspect of their work.. However, it is unrealistic to expect teachers and pupils' to practice such separation. The teachers ignored this advice and worked to link formative and summative productively. This section sets out ways in which this was done: However, these approaches were only used for tests where teachers had control over the setting and the marking; their application was more limited for tests where the teachers had little or no control.

Overall, the main possibilities for improving classroom practice are as follows:

- Pupils should be engaged in a reflective review of the work they have done to enable them to plan their revision effectively.

- Pupils should be encouraged to set questions and mark answers to help them, both to understand the assessment process and to focus further efforts for improvement.

- Pupils should be encouraged through peer- and self-assessment to apply criteria to help them understand how their work might be improved. This may include providing opportunities for pupils to rework examination answers in class.

The main overall message is that summative tests should be, and should be seen to be, a positive part of the learning process. By active involvement in the test process, pupils can see that they can be beneficiaries rather than victims of testing, because tests can help them improve their learning.

Reflections: teachers, students and learning.

In this section we reflect on the deeper issues about learning and about teaching that have been provoked, not least amongst the teachers involved, by the changes that are entailed by improved assessment for learning.

The big idea: 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 pupils feedback that they knew in advance was going to be useful. To do that they needed to build up models of how pupils learn.

So the teachers came to take greater care in selecting tasks, questions and other prompts to ensure that the responses made by pupils actually helped the teaching process. Such responses can 'put on the table' the ideas which pupils bring to a learning task. The key to effective learning is to then find ways to help pupils 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 pupils' 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 principle of the constructivist view of learning – to start where the pupils is and to involve the pupils 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 pupil; it has to be done by the pupil.

Pupils 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 pupils using criteria to assess the work of their peers. The activities, by encouraging pupils 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 pupils 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 pupils would become more salient and effective.

A learning environment: changes of role

There are also deeper issues here. A learning environment has to be 'engineered' to involve pupils' more actively in the tasks. The emphasis has to be on the pupils 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 pupils 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 one of sharing responsibility for the class's learning with the class—exactly the same process, but viewed from two very different perspectives. In one perspective, the teachers and pupils are in a delivery-recipient relationship, in the other they are partners in pursuit of a shared goal:

> *What formative assessment has done for me is made me focus less on myself but more on the children. I have had the confidence to empower the students to take it forward*
>
> *Robert, Two Bishops' School*

The learning environment envisaged requires a classroom culture that may well be unfamiliar and disconcerting for both teachers and pupils. 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 pupils. As Perrenoud (1991) put it:

> *Every teacher who wants to practice formative assessment must reconstruct the teaching contracts so as to counteract the habits acquired by his pupils"*

For the pupils, 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 pupils 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?'

What has been happening here is that everybody's role expectations, i.e. what teachers and pupils think that being a teacher or being a pupil 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.

To summarise, expectations and classroom culture can be changed by:

- Incorporating the changes in the teacher's role one step at a time, as they seem appropriate.
- Empowering pupils to become active learners, taking responsibility for their own learning
- Changing the 'classroom contract' so that all expect that teacher and pupils work together for the same end, the improvement of everyone's learning.
- Sustained attention to, and reflection on, assessment for learning issues.

Motivation and self-esteem

Learning is not just a cognitive exercise. The need to motivate pupils is evident, but it is often assumed that this is best done by offering such extrinsic rewards as merits, grades, gold stars and prizes. There is ample evidence that challenges this assumption.

Pupils will only invest effort in a task if they believe that they can achieve something. If a learning exercise is seen as a competition, then everyone is aware that there will be losers as well as winners: those who have a track record as losers will see little point in trying. Thus, the problem is to motivate everyone, even although some are bound to achieve less than others. In tackling this problem, the type of feedback given is very important. Many research studies support this assertion. Examples are:

- Pupils told that feedback ' …will help you to learn' learn more than those told that 'how you do tells us how smart you are and what grades you'll get'; the difference is greatest for low attainers (Newman & Schwager, 1995).

- Those given feedback as marks are likely to see it as a way to compare themselves with others (ego-involvement), those given only comments see it as helping them to improve (task-involvement): the latter group out-perform the former (Butler, 1987).

- In a competitive system, low attainers attribute their performance to lack of 'ability', high attainers to their effort; in a task oriented system, all attribute to effort, and learning is improved, particularly amongst low attainers (Craven *et al.* 1991).

- A comprehensive review of research studies of feedback showed that feedback improved performance in 60% of them. In the cases where it was not helpful, the feedback turned out to be merely a judgment or grading with no indication of how to improve (Kluger & DeNisi, 1996).

In general, feedback given as rewards or grades enhances ego rather than task involvement. It can focus pupils' attention on their 'ability' rather than on the importance of effort, damaging the self-esteem of low attainers and leading to problems of 'learned helplessness' (Dweck 1986, 2000). Feedback that focuses on what needs to be done can encourage all to believe that they can improve. Such feedback can enhance learning, both directly through the effort that can ensue, and indirectly by supporting the motivation to invest such effort.

Whilst the work just quoted refers mainly to day-to-day classroom work, the negative effects of ego-oriented feedback are even more striking when high-stakes tests are involved. Their damaging effects of high stakes tests on pupils' motivation and learning work have been strikingly illustrated by a recent survey of research on testing, motivation and learning published by the UK Assessment Reform Group (ARG 2002). A widespread search found 183 potentially relevant research studies of which 19 were selected, by application of rigorous criteria of quality and relevance, as providing sound and valid empirical evidence. Some of the main findings were :

- The use of repeated practice tests emphasises the importance of the tests and encourages pupils to adopt test-taking strategies designed to avoid effort and responsibility. Repeated practice tests are thereby detrimental to higher order thinking.

- When tests pervade the ethos of the classroom, test performance is more highly valued than what is being learned. Yet pupils, even at primary school, are aware that tests give only a narrow view of their learning.

- These tests have a significant impact on self-esteem. After the introduction of the National Curriculum tests in England, low-achieving students had lower self-esteem than higher achieving students. Before the tests were introduced there was no correlation between self-esteem and achievement.

- Lower achieving students are doubly disadvantaged by summative assessment. Being labelled as failures has an impact on how they feel about their ability to learn. It also lowers further their already low self-esteem and reduces the chance of future effort and success .

Summative tests

Some have argued that formative and summative assessments are so different in their purpose that they have to be kept apart, and such arguments are strengthened by experience of the harmful influence that narrow 'high-stakes' summative tests can have on teaching. However, it is unrealistic to expect teachers and pupils' to practice such separation, so the challenge is to achieve a more positive relationship between the two.

In fact, as our project showed, classroom practice in relation to preparation for summative testing can change, by giving pupils help to engage in a reflective review of the work they have done so that they can plan their revision effectively, by asking them to set questions, and by encouraging them to mark answers through peer- and self-assessment. Such changes help pupils to grasp the criteria of quality and so to understand how their work might be improved. They also help them both to understand the assessment process and to gain an overview of their work.

The underlying message is that summative tests can, and should be seen to be, a positive part of the learning process. By active involvement in the test process, pupils can see that they can be beneficiaries rather than victims of testing, because tests can help them improve their learning.

Teachers in the formative-summative interface

However, the benefits of the formative approach could only be realised for tests over which the teachers had full control. Matters looked quite different when our project teachers were asked to write about and discuss the interface between their formative work and external test pressures. When asked how summative assessments affect what they taught and the way in which they taught it, the science teachers in the project felt that they had to teach to the test, although the test questions seemed to them to lack validity in relation to the national curriculum specifications. These summative pressures inhibited teaching to cross-link concepts – everything had to be compartmentalised, isolated. The test pressures also limited practical work, affected the range of types of lesson, and restricted both their teaching styles and pupil involvement and enjoyment. Teachers of English believed that training pupils to write in the timed conditions of external 'high-stakes' tests absorbed valuable time, and this

specific training was not necessarily connected to their view of valid English teaching. They believed there was 'something desiccated about the process'.

The effects were pervasive. Thus although teachers had become more sensitive and more expert in the framing of productive questions for everyday use, there was very little use of questions set by the teachers themselves in their school tests. They relied on questions taken from past external tests and from text-books, and often the same test was used year after year.

In marking**,** many said they would like to produce a profile, but that it involved too much work and they did not have the time. Whilst they were well aware that aggregation of test marks loses all the useful information for feedback, they nevertheless provided such data "for the school office". These practices use did not go with satisfaction with the external sources and procedures. Some believed that their own tests were more valid, but felt that they would not have external credibility, whether with colleagues or with the pupils.

School requirements for formal test occasions were very variable. In frequency they ranged from none to four or five times year, at fixed times or at variable times. It was common to have a formal test at the end of a topic or module; some tested before the end of a module so that there would be time for formative use of the results. One teacher had a firm belief in testing topic by topic and avoiding any overall 'big bang' test. Another teacher would only set a test when she thought the class ready – if the homework showed that they were not, she would defer the test and go over the common difficulties.

A few believed that they only set tests because their schools required them – as one teacher put it :

> *I know a lot more about this class because of the formative assessment . I mean we discuss things, they do presentations, they talk to me, I talk to them, they talk to each other – and I could tell you for every one of this class their strengths and weaknesses.*

However, some holding this belief nevertheless realised also that their staff colleagues did not know the pupils in their classes well enough to assess each of them using personal knowledge. In addition, they were concerned that this variability in assessment expertise was exacerbated by the flux of teacher appointments and resignations, and by the frequent use of substitute teachers. Thus the prospects of reliance on formative assessment as sole basis for review across all teachers in their subject department seemed to them to be remote.

Improved formative practices can generate a lot of data about pupils' progress, and it would seem that a further step might be to record and accumulate this to strengthen teachers' contribution to summative reporting. However, it seemed that teachers were not able to exploit this potential for synergy. In the practices of recording and reporting it was again hard to find any regularities. For homework, some kept no record, some recorded only that the homework has been attempted, some marked but recorded the mark only in their own record book, some used peer assessment, and perhaps checked this and recorded a mark afterwards. For a few, they could envisage that a complete record could involve a log of homework results, plus end-of-module tests scores, plus scores on less formal class tests, plus effort grades judged by the teachers (or in one case, by the pupils). One school set up

software so that pupils could enter their own records, up-date them, and review progress themselves and with their teacher. A few said that it was too hard to review all the data – so they just used the scores on a terminal test to report to the school management, particularly when the need arose at the end of term when they did not have the time to collect and analyse the data.

Another aspect of external pressures on assessment was the need to report to parents. Here, two contrary views were expressed. One teacher explained how he felt he had to use only the test score:

> *. . . . everyone is obsessed with numbers rather than what the kids really know. I think parents don't put a lot of weighting onto teacher assessment, they don't even put a lot of weighting onto coursework as a way of measuring their child's ability. Certainly they wouldn't put a lot of store by what I say as opposed to what a test says because at the end of the day society is driven by test results in schools . . at the end of the day if you gave them the choice – they can have what I think or they can have an exam grade – they would take the exam grade every time because that is what they want. They want to be able to compare that with everyone else even though it is not about everyone else.*

Given such perception of parental expectations, the need to avoid argument with parents, and the shortages of time, use only of a score on a terminal test was the easy way out.

Other teachers expressed a contrary view contrary view. Their experience was that the rich detail provided through good formative assessment was appreciated by parents, in that they and their children could be given specific and helpful advice on how to improve. This different appraisal of the situation may have arisen from a variety of causes, to do with the success of the school as a whole in explaining learning needs to parents.

.At the end of one discussion of these issues with a group of the project teachers, one of them summed it up as follows:

> *It is a bit depressing that isn't it?*

We could only agree with her.

Conclusion

Even with teachers who had achieved so much in changing the quality and effectiveness of their work, the reality of the pressures of external tests had to be acknowledged and accommodated despite their view of the low quality and poor effects on learning of such tests. Overall teachers seemed to be trapped in a 'no-man's land', between their new commitment to formative assessment and the different, often contradictory, demands of the external test system. Their formative use of summative tests had enabled to move the frontier significantly, but further territory seemed un-assailable.

The message of this paper is that there is extensive evidence that enhancement of formative assessment can raise standards of learning and develop teachers' professionalism in way that they value. However, the full potential here is blocked by the present regimes of external testing, so that unless these regimes are reformed to create a new synergy between teachers'

and the public to satisfy summative needs, standards cannot be raised. That this can be done has been established elsewhere: the notable example is the state of Queensland, Australia, where external testing for school leaving certificates and university entrance was abolished in 1982 and all needs are met by teachers' assessments, albeit backed by a rigorous system of inter-calibration between schools through meetings in regional clusters (Butler, 1995).

The possibility of giving more weight to teachers' summative assessments must however be looked at with care and caution and the Queensland initiative reported above required a few years of development to attain adequate quality. A detailed review of research on teachers' summative assessments has shown several examples of innovations that failed because the results did not meet the standards on reliability and validity that anyone would require for high-stakes test results (Harlen, 2004). This there is need for careful trial work to see how high quality can be attained before any new initiatives could be taken in this area.

References

ARG (2002) *Testing, Motivation and Learning* University of Cambridge School of Education : Assessment Reform Group

Black, P. & Wiliam, D (1998a) Assessment and Classroom Learning. *Assessment in Education,* **5**(1), 7-71.

Black, P. & Wiliam, D. (1998b) *Inside the Black Box: raising standards through classroom assessment* London: School of Education, King's College.

Black, P. J.; Harrison, C.; Lee, C.; Marshall, B. & Wiliam, D (2002) *Working inside the black box: Assessment for learning in the classroom*. London, UK: King's College London School of Education.

Black, P.; Harrison, C.; Lee, C.; Marshall, B. & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Buckingham, UK: Open University Press.

Black, P. and Wiliam, D. (2003) 'In Praise of Educational Research': formative assessment. British Educational Research Journal. **29**(5), 623-37.

Butler, R. (1987). Task-Involving and Ego-Involving Properties of Evaluation: Effects of Different Feedback Conditions on Motivational Perceptions, Interest and Performance. *Journal of Educational Psychology,* **79** (4), 474-482.

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.

Butler, J. (1995). Teachers Judging Standards in Senior Science Subjects: Fifteen Years of the Queensland Experiment. *Studies in Science Education,* **26**, pp. 135-157.

Craven, R. G., Marsh, H. W. & Debus, R. L. (1991). Effects of Internally Focused Feedback on Enhancement of Academic Self-Concept. *Journal of Educational Psychology,* **83** (1), 17-27.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist (Special Issue: Psychological science and education),* **41** (10), 1040-1048.

Dweck, C. S. (2000). *Self-theories:Their role in motivation, personality and development.* London: Taylor and Francis.

Foos, P. W., Mora, J. J. & Tkacz, S. (1994). Student Study Techniques and the Generation Effect. *Journal of Educational Psychology,* **86** (4), 567-576.

Harlen, W. (2004) *A systematic review of the evidence of the reliability and validity of assessment by teachers for summative purposes*. London: EPPI Centre Institute of Education.

King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist,* **27** (1), 111-126.

Kluger, A. N. & 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), 254-284.

Newman, R. S. & Schwager, M. T. (1995). Students' Help Seeking During Problem Solving: Effects of Grade, Goal, and Prior Achievement. *American Educational Research Journal,* **32** (2), 352-376.

Perrenoud, P. (1991). Towards a pragmatic approach to formative evaluation, in: P. Weston (Ed.) *Assessment of Pupils Achievement: Motivation and School Success,* pp. 79—101. (Amsterdam: Swets and Zeitlinger).

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, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science,* **18**, 119-144.

Sadler, R. (1998). Formative assessment: revisiting the territory. *Assessment in Education* **5**(1), 77-84

Wiliam, D., Lee, C., Harrison, C. and Black, P. (2004) Teachers developing assessment for learning: impact on student achievement. *Assessment in Education,* **11**(1), 49-65.