

Working Inside the Black Box: Assessment for Learning in the Classroom.

by Paul Black, Christine Harrison, Clare Lee, Bethan Marshall and Dylan Wiliam

© COPYRIGHT 2004 Phi Delta Kappa, Inc.

In their widely read article "Inside the Black Box," Mr. Black and Mr. Wiliam demonstrated that improving formative assessment raises student achievement. Now they and their colleagues report on a follow-up project that has helped teachers change their practice and students change their behavior so that everyone shares responsibility for the students' learning.

IN 1998 "Inside the Black Box," the predecessor of this article, appeared in this journal.¹ Since then we have learned a great deal about the practical steps needed to meet the purpose expressed in the article's subtitle: "raising standards through classroom assessment."

In the first part of "Inside the Black Box," we set out to answer three questions. The first was, Is there evidence that improving formative assessment raises standards? The answer was an unequivocal yes, a conclusion based on a review of evidence published in over 250 articles by researchers from several countries.² Few initiatives in education have had such a strong body of evidence to support a claim to raise standards.

This positive answer led naturally to the second question: Is there evidence that there is room for improvement? Here again, the available evidence gave a clear and positive answer, presenting a detailed picture that identified three main problems: 1) the assessment methods that teachers use are not effective in promoting good learning, 2) grading practices tend to emphasize competition rather than personal improvement, and 3) assessment feedback often has a negative impact, particularly on low-achieving students, who are led to believe that they lack "ability" and so are not able to learn.

However, for the third question -- Is there evidence about how to improve formative assessment? -- the answer was less clear. While the evidence provided many ideas for improvement, it lacked the detail that would enable teachers to implement those ideas in their classrooms. We argued that teachers needed "a variety of living examples of implementation."

The Journey: Learning with Teachers

Since 1998, we have planned and implemented several programs in which groups of teachers in England have been supported in developing innovative practices in their classrooms, drawing on the ideas in the original article. While this effort has amply confirmed the original proposals, it has also added a wealth of new findings that

are both practical and authentic. Thus we are now confident that we can set out sound advice for the improvement of classroom assessment.

The KMOFAP Project

To carry out the exploratory work that was called for, we needed to collaborate with a group of teachers willing to take on the risks and extra work involved, and we needed to secure support from their schools and districts. Funding for the project was provided through the generosity of the Nuffield Foundation, and we were fortunate to find two school districts -- Oxfordshire and Medway, both in southern England -- whose supervisory staff members understood the issues and were willing to work with us. Each district selected three secondary schools: Oxfordshire chose three coeducational schools, and Medway chose one coeducational school, one boys' school, and one girls' school. Each school selected two science teachers and two mathematics teachers. We discussed the plans with the principal of each school, and then we called the first meeting of the 24 teachers. So in January 1999, the King's-Medway-Oxfordshire Formative Assessment Project (KMOFAP) was born.

Full details of the project can be found in our book, *Assessment for Learning: Putting It into Practice*.³ 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 College team, records of meetings of the whole group of teachers, interviews with and writing by the teachers themselves, and a few discussions with student groups. Initially, we worked with science and mathematics teachers, but the work has been extended more recently to involve teachers of English in the same schools and teachers of other subjects in other schools.

Spreading the Word

Throughout the development of the project, we have responded to numerous invitations to talk to other groups of teachers and advisers. Indeed, over five years we have made more than 400 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 gives us confidence that our general findings will be of value to all, although some important details may differ for different age groups and subjects. Furthermore, a group at Stanford University obtained funding from the National Science Foundation to set up a similar development project, in collaboration with King's, in schools in California.

Working Inside the Black Box: Assessment for Learning in the Classroom.

Extension of our own work has been made possible by this funding. And we also acknowledge support from individuals in several government agencies who sat on the project's steering group, offered advice and guidance, and helped ensure that assessment for learning (see "Assessment for Learning," below) is a central theme in education policy in England and Scotland.

The Learning Gains

From our review of the international research literature, we were convinced that enhanced formative assessment would produce gains in student achievement, even when measured in such narrow terms as scores on state-mandated tests. At the outset we were clear that it was important to have some indication of the kinds of gains that could be achieved in real classrooms and over an extended period of time. Since each teacher in the project was free to choose the class that would work on these ideas, we discussed with each teacher what data were available within the school, and we set up a "mini-experiment" for each teacher.

Each teacher decided what was to be the "output" measure for his or her class. For grade-10 classes, this was generally the grade achieved on the national school-leaving examination taken when students are 16 (the General Certificate of Secondary Education or GCSE). For grade-8 classes, it was generally the score or level achieved on the national tests administered to all 14-year-olds. For other classes, a variety of measures were used, including end-of-module-test scores and marks on the school's end-of-year examinations.

For each project class, the teacher identified a comparison 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 or, failing that, a nonparallel class taught by the same or a different teacher. When the project and the control classes were not strictly parallel, we controlled for possible differences in prior achievement by the use of "input" measures, such as school test scores from the previous year or other measures of aptitude.

This approach meant that the size of the improvement was measured differently for each teacher. For example, a grade-10 project class might outperform the comparison class by half a GCSE grade, but another teacher's grade-8 project class might outscore its control class by 7% on an end-of-year exam. To enable us to aggregate the results, we adopted the common measuring stick of the "standardized effect size," calculated by taking the difference between the scores of the experimental and control groups and then dividing this number by the

standard deviation (a measure of the spread in the scores of the groups).

For the 19 teachers on whom we had complete data, the average effect size was around 0.3 standard deviations. Such improvements, produced across a school, would raise a school in the lower quartile of the national performance tables to well above average. Thus it is clear that, far from having to choose between teaching well and getting good test scores, teachers can actually improve their students' results by working with the ideas we present here.

How Change Can Happen

We set out our main findings about classroom work under four headings: questioning, feedback through grading, peer- and self-assessment, and the formative use of summative tests. Most of the quotations in the following pages are taken directly from pieces written by the teachers. The names of the teachers and of the schools are pseudonyms, in keeping with our policy of guaranteeing anonymity.

Questioning

Many teachers do not plan and conduct classroom dialogue in ways that might help students to learn. Research has shown that, after asking a question, many teachers wait less than one second and then, if no answer is forthcoming, ask another question or answer the question themselves.⁴ A consequence of such short "wait time" is that the only questions that "work" are those that can be answered quickly, without thought -- that is, questions calling for memorized facts. Consequently, the dialogue is at a superficial level. As one teacher put it:

I'd become dissatisfied with the closed Q & A style that my unthinking teaching had fallen into, and I would frequently be lazy in my acceptance of right answers and sometimes even tacit complicity with a class to make sure none of us had to work too hard. . . . They and I knew that if the Q & A wasn't going smoothly, I'd change the question, answer it myself, or only seek answers from the "brighter students." There must have been times (still are?) where an outside observer would see my lessons as a small discussion group surrounded by many sleepy onlookers. -- James, Two Bishops School

The key to changing such a situation is to allow longer wait time. But many teachers find it hard to do this, for it requires them to break their established habits. Once they change, the expectations of their students are challenged:

Increasing waiting time after asking questions proved

Working Inside the Black Box: Assessment for Learning in the Classroom.

difficult to start with due to my habitual desire to "add" something almost immediately after asking the original question. The pause after asking the question was sometimes "painful." It felt unnatural to have such a seemingly "dead" period, but I persevered. Given more thinking time, students seemed to realize that a more thoughtful answer was required. Now, after many months of changing my style of questioning, I have noticed that most students will give an answer and an explanation (where necessary) without additional prompting. -- Derek, Century Island School

One teacher summarized 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 are given time to explore answers together. My year 8 [grade 7] 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 help more students become involved in discussions and increase the length of their replies. Another way to broaden participation is to ask students to brainstorm ideas, perhaps in pairs, for two to three minutes before the teacher asks for contributions. Overall, a consequence of such changes is that teachers learn more about the students' prior knowledge and about any gaps and misconceptions in that knowledge, so that teachers' next moves can better address the learners' real needs.

To exploit such changes means moving away from the routine of limited factual questions and refocusing attention on the quality and the different functions of classroom questions. Consider, for example, the use of a "big question": an open question or a problem-solving task that

can set the scene for a lesson and evoke broad discussion or prompt focused small-group discussions. However, if this strategy is to be productive, both the responses that the task might generate and the ways of following up on these responses have to be anticipated. Collaboration between teachers to exchange ideas and experiences about good questions is very valuable. The questions themselves then become a more significant part of teaching, with attention focused on how they can be constructed and used to explore and then develop students' learning. Here's one teacher's thinking on the matter:

I chose a year-8, middle-band group and really started to think about the type of questions I was asking -- were they just instant one-word answers -- what were they testing -- knowledge or understanding -- was I giving the class enough time to answer the question, was I quickly accepting the correct answer, was I asking the girl to explain her answer, how was I dealing with a wrong answer? When I really stopped to think, I realized 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

Effective questioning is also an important aspect of the impromptu interventions teachers conduct once the students are engaged in an activity. Asking simple questions, such as "Why do you think that?" or "How might you express that?" can become part of the interactive dynamic of the classroom and can provide an invaluable opportunity to extend students' thinking through immediate feedback on their work.

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, that is, questions that explore issues that are critical to the development of students' understanding.

* Wait time has to be increased to several seconds in order to give students 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. The aim is thoughtful improvement rather than getting it right the first time.

* Follow-up activities have to be rich, in that they create opportunities to extend students' understanding.

Put simply, the only point of asking questions is to raise issues about which a teacher needs information or about which the students need to think. When such changes have been made, experience demonstrates that students

Working Inside the Black Box: Assessment for Learning in the Classroom.

become more active participants and come to realize 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. The teachers also shift in their role, from presenters of content to leaders of an exploration and development of ideas in which all students are involved.

Feedback Through Grading

When giving students feedback on both oral and written work, it is the nature, rather than the amount, of commentary that is critical. Research experiments have established that, while student learning can be advanced by feedback through comments, the giving of numerical scores or grades has a negative effect, in that students ignore comments when marks are also given.⁵ These results often surprise teachers, but those who have abandoned the giving of marks discover that their experience confirms the findings: students do engage more productively in improving their work.

Many teachers will be concerned about the effect of returning students' work with comments but no scores or grades. There may be 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 [Lee] observed on several occasions how little time pupils spent reading my comments if there were grades given as well. My routine is now, in my target class, i) to not give grades, only comments; ii) to give comments that highlight what has been done well and what needs further work; and iii) to give the minimum follow-up work expected to be completed next time I mark the books. -- Nancy, Riverside School

Initial fears about how students might react turned out to be unjustified, and neither parents nor school inspectors have reacted adversely. Indeed, the provision of comments to students helps parents to focus on the learning issues rather than on trying to interpret a score or grade. We now believe that the effort that many teachers devote to grading homework may be misdirected. A numerical score or a grade does not tell students how to improve their work, so an opportunity to enhance their learning is lost.

A commitment to improve comments requires more work initially, as teachers have to attend to the quality of the comments that they write on students' work. Collaboration between teachers in sharing examples of effective comments can be very helpful, and experience will lead to

more fluency. There is, however, more involved because comments become useful feedback only if students use them to guide further work, so new procedures are needed.

After the first INSET [inservice training meeting] I was keen to try out a different way of marking books to give pupils more constructive feedback. I was keen to try and have a more easy method of monitoring pupils' response to my comments without having to trawl through their books each time to find out if they'd addressed my comments. I implemented a comment sheet at the back of my year-8 class' books. It is A4 [letter] in size, and the left-hand side is for my comments, and the right-hand side is for the pupils to demonstrate by a reference to the page in their books where I can find the evidence to say whether they have done the work. . . . The comments have become more meaningful as the time has gone on, and the books still take me only one hour to mark. -- Sian, Cornbury Estate School

We have encountered a variety of ways of accommodating the new emphasis on comments. Some teachers have ceased assigning scores or grades at all, some teachers enter scores in their own record books but do not write them in the students' books, others give a score or grade only after a student has responded to the teacher's comments. Some teachers spend more time on certain pieces of work to ensure that they obtain good feedback and, to make time for this, either do not mark some pieces, or look at only a third of their students' books each week, or involve the students in checking the straightforward tasks.

A particularly valuable method is to devote some lesson time to rewriting selected pieces of work, so that emphasis can be put on feedback for improvement within a supportive environment. This practice can change students' expectations about the purposes of class work and homework.

As they tried to create useful feedback comments, many of the project teachers realized that they needed to reassess the work that they had asked students to undertake. They found that some tasks were useful in revealing students' understandings and misunderstandings, while 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 of feedback can be summarized as follows:

* Written tasks, alongside oral questioning, should encourage students to develop and show understanding of

Working Inside the Black Box: Assessment for Learning in the Classroom.

the key features of what they have learned.

* Comments should identify what has been done well and what still needs improvement and give guidance on how to make that improvement.

* Opportunities for students to respond to 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. The implementation of such reforms can change both teachers' and students' attitudes toward written work: the assessment of students' work will be seen less as a competitive and summative judgment and more as a distinctive step in the process of learning.

Peer Assessment and Self-Assessment

Students can achieve a learning goal only if they understand that goal and can assess what they need to do to reach it. So self-assessment is essential to learning.⁶ Many teachers who have tried to develop their students' self-assessment skills have found that the first and most difficult task is to get students 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 that allows them to manage and control it for themselves. In other words, students are developing the capacity to work at a metacognitive level.

In practice, peer assessment turns out to be an important complement to self-assessment. Peer assessment is uniquely valuable because students may accept criticisms of their work from one another that they would not take seriously if the remarks were offered by a teacher. Peer work is also valuable because the interchange will be in language that students themselves naturally use and because students learn by taking the roles of teachers and examiners of others.⁷ One teacher shared her positive views of peer assessment:

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 led 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 teacher's comments brings out an important point: when students do not understand an explanation, they are likely to interrupt a fellow student 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 students. The teacher can be free to observe and reflect on what is happening and to frame helpful interventions:

We regularly do peer marking -- I find this very helpful indeed. A lot of misconceptions come to the fore, and we then discuss these as we are going over the homework. I then go over the peer marking and talk to pupils individually as I go round the room. -- Rose, Brownfields School

However, self-assessment will happen only if teachers help their students, particularly the low achievers, to develop the skill. This can take time and practice:

The kids are not skilled in what I am trying to get them to do. I think the process is more effective long term. If you invest time in it, it will pay off big dividends, this process of getting the students to be more independent in the way that they learn and to take the responsibility themselves. -- Tom, Riverside School

One simple and effective idea is for students to use "traffic light" icons, labeling their work green, yellow, or red according to whether they think they have good, partial, or little understanding. These labels serve as a simple means of communicating students' self-assessments. Students may then be asked to justify their judgments in a peer group, thus linking peer assessment and self-assessment. This linkage can help them develop the skills and the detachment needed for effective self-assessment.

Another approach is to ask students first to use their "traffic light" icons on a piece of work and then to indicate by hands-up whether they put a green, yellow, or red icon on it. The teacher can then pair up the greens and the yellows to help one another deal with their problems, while the red students meet with the teacher as a group to deal with their deeper problems. For such peer-group work to succeed, many students will need guidance about how to behave in groups, including such skills as listening to one another and taking turns.

In some subjects, taking time to help students understand scoring rubrics is also very helpful. Students can be given simplified versions of the rubrics teachers use, or they can be encouraged to rewrite them or even to create their own. Again, peer assessment and self-assessment are intimately linked. Observers in several language arts classrooms saw children apply to their own work lessons

Working Inside the Black Box: Assessment for Learning in the Classroom.

they had learned in peer assessment. A frequently heard comment was "I didn't do that either" or "I need to do that too."

Students' reflection about their understanding can also be used to inform future teaching, and their feedback can indicate in which areas a teacher needs to spend more time. A useful guide is to ask students to "traffic light" an end-of-unit test at the beginning of the unit: the yellow and red items can be used to adjust priorities within the teaching plan. Our experience leads us to offer the following recommendations for improving classroom practice:

- * 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. Such criteria may well be abstract, but concrete examples should be used in modeling exercises to develop understanding.

- * Students should be taught the habits and skills of collaboration in peer assessment, both because these are of intrinsic value and because peer assessment can help develop the objectivity required for effective self-assessment.

- * Students should be encouraged to keep in mind the aims of their work and to assess their own progress toward meeting these aims as they proceed. Then they will be able to guide their own work and so become independent learners.

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

The Formative Use of Summative Tests

The practices of self-assessment and peer assessment can be applied to help students prepare for tests, as in tackling the following problem:

[The students] 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 [reviewing] 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 remedy this situation, students can be asked to "traffic light" a list of key words or the topics on which the test will

be set. The point of this exercise is to stimulate the students to reflect on where they feel their learning is secure, which they mark green, and where they need to concentrate their efforts, in yellow and red. These traffic lights then form the basis of a review plan. Students can be asked to identify questions on past tests that probe their "red" areas. Then they can work with textbooks and in peer groups to ensure that they can successfully answer those questions.

The aftermath of tests can also be an occasion for formative work. Peer marking of test papers can be helpful, as with normal written work, and it is particularly useful if students are required first to formulate a scoring rubric -- an exercise that focuses attention on the criteria of quality relevant to their productions. After peer marking, teachers can reserve their time for discussion of the questions that give widespread difficulty, while peer tutoring can tackle those problems encountered by only a minority of students.

One other finding that has emerged from research studies is that students trained to prepare for examinations by generating and then answering their own questions outperformed comparable groups who prepared in conventional ways.⁸ Preparing test questions helps students develop an overview of the topic:

Pupils have had to think about what makes a good question for a test and in doing so need to have a clear understanding of the subject material. As a development of this, the best questions have been used for class tests. In this way, the pupils can see that their work is valued, and I can make an assessment of the progress made in these areas. When going over the test, good use can be made of group work and discussions between students concentrating on specific areas of concern. -- Angela, Cornbury Estate School

Developments such as these 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 when one experiences the harmful influence that narrow, high-stakes summative tests can have on teaching. However, it is unrealistic to expect teachers and students to practice such separation, so the challenge is to achieve a more positive relationship between the two. All of the ways we have described for doing so can be used for tests in which teachers have control over the setting and the marking. But their application may be more limited for tests in which the teacher has little or no control.

Overall, the main possibilities for improving classroom practice by using summative tests for formative purposes

Working Inside the Black Box: Assessment for Learning in the Classroom.

are as follows:

- * Students can be engaged in a reflective review of the work they have done to enable them to plan their revision effectively.
- * Students can be encouraged to set questions and mark answers so as to gain an understanding of the assessment process and further refine their efforts for improvement.
- * Students should be encouraged through peer assessment and self-assessment to apply criteria to help them understand how their work might be improved. This may include providing opportunities for students to rework examination answers in class.

The overall message is that summative tests should become a positive part of the learning process. Through active involvement in the testing process, students can see that they can be the beneficiaries rather than the victims of testing, because tests can help them improve their learning.

Reflections: Some Underlying Issues

The changes that are entailed by improved assessment for learning have provoked us and the teachers involved to reflect on deeper issues about learning and teaching.

Learning Theory

One of the most surprising things that happened during the early INSET sessions was that the participating teachers asked us to run a session on the psychology of learning. In retrospect, perhaps we should not have been so surprised at this request. After all, we had stressed that feedback functioned formatively only if the information fed back to the learner was used by the learner in improving performance. But while one can work out after the fact whether or not any feedback has had the desired effect, what the teachers needed was a way 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 students' responses actually helped the teaching process. Such responses can "put on the table" the ideas that students bring to a learning task. The key to effective learning is then to 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 was not a

process of passive reception of knowledge, but one in which the learners were active in creating their own understandings. Put simply, it became clear that, no matter what the pressure to achieve good test scores, learning must be done by the student.

Students came to understand what counted as good work through exemplification. Sometimes this was done through focused whole-class discussion around a particular example; at other times it was achieved through the use of sets of criteria to assess the work of peers.

Engaging in peer assessment and self-assessment is much more than just checking for errors or weaknesses. It involves making explicit what is normally implicit, and thus it requires students to be active in their learning. As one student wrote:

After a pupil marking my investigation, I can now acknowledge my mistakes easier. I hope that it is not just me who learned from the investigation but the pupil who marked it did also. Next time I will have to make my explanations clearer, as they said "It is hard to understand." . . . I will now explain my equation again so it is clear.

The students also became much more aware of when they were learning and when they were not. One class, which was 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?" While students who are in tune with their learning can create difficulties for teachers, we believe that these are exactly the kinds of problems we should want to have.

Subject Differences

From hearing about research and discussing ideas with other colleagues, the teachers built up a repertoire of generic skills. They planned their questions, allowed appropriate wait time, and gave feedback that was designed to cause thinking. They ensured that students were given enough time during lessons to evaluate their own work and that of others.

However, after a while it became clear that these generic strategies could go only so far. Choosing a good question requires a detailed knowledge of the subject, but not necessarily the knowledge that is gained from advanced study in a subject. A high level of qualification in a subject is less important than a thorough understanding of its fundamental principles, an understanding of the kinds of difficulties that students might have, and the creativity to be able to think up questions that stimulate productive

Working Inside the Black Box: Assessment for Learning in the Classroom.

thinking.⁹ Furthermore, such pedagogical content knowledge is essential in interpreting responses. That is, what students say will contain clues to aspects of their thinking that may require attention, but picking up on these clues requires a thorough knowledge of common difficulties in learning the subject. Thus, while the general principles of formative assessment apply across all subjects, the ways in which they manifest themselves in different subjects may differ. We have encountered such differences in making comparisons between teachers of mathematics, science, and language arts.

In mathematics, students have to learn to use valid procedures and to understand the concepts that underpin them. Difficulties can arise when students learn strategies that apply only in limited contexts and do not realize that they are inadequate elsewhere. Questioning must then be designed to bring out these strategies for discussion and to explore problems in understanding the concepts so that students can grasp the need to change their thinking. In such learning, there is usually a well-defined correct outcome. In more open-ended exercises, as in investigations of the application of mathematical thinking to everyday problems, there may be a variety of good solutions. Then an understanding of the criteria of quality is harder to achieve and may require joint discussion of examples and of the abstract criteria that they exemplify.

In science, the situation is very similar. There are many features of the natural world for which science provides a "correct" model or explanation. However, outside school, many students acquire different ideas. For example, some students come to believe that animals are living because they move but that trees and flowers are not because they don't. Or students may believe that astronauts seem almost weightless on the moon because there is no air present. Many of these "alternative conceptions" can be anticipated, for they have been well documented. What has also been documented is that the mere presentation of the "correct" view has been shown to be ineffective. The task in such cases is to open up discussion of such ideas and then provide feedback that challenges them by introducing new pieces of evidence and argument that support the scientific model.

There are other aspects for which an acceptable outcome is less well defined. As in mathematics, open-ended investigations call for different approaches to formative assessment. Even more open are issues about social or ethical implications of scientific achievements, for there is no "answer." Thus such work has to be "open" in a more fundamental way. Then the priority in giving feedback is to challenge students to tease out their assumptions and to help them to be critical about the quality of any arguments.

Peer assessment and self-assessment have a long history in language arts. Both the nature of the subject and the open outcome of many of the tasks characteristically make such practices central to one of the overall aims of the discipline, which is to enhance the critical judgment of the students.

A second important function of peer assessment and self-assessment was introduced by Royce Sadler, who argued that criteria alone are unhelpful in judging the quality of a piece of work or in guiding progression, because there will always be too many variables.¹⁰ The key lies in knowing how to interpret the criteria in any particular case, which involves "guild knowledge." Teachers acquire this knowledge through assessing student work, and it is this process that allows them to differentiate between grades and to gain a sense of how progress is achieved. Peer assessment and self-assessment provide similar opportunities for students to be apprenticed into the guild, provided the criteria of quality are clearly communicated.

In language arts, as in science and mathematics, attention needs to be paid to the central activities. Those that are the most successful are those rich tasks that provide students with an opportunity either to extend their understanding of a concept within the text or to "scaffold" their ideas before writing. Characteristically, these include small- group and pair work, with the results often being fed back into a whole- class discussion. Again, this type of work is not uncommon in language arts, the skill being to make the task sufficiently structured to scaffold learning but not so tightly defined as to limit thinking. Such activities not only provide students with a chance to develop their understanding through talk, but they also provide the teacher with the opportunity to give feedback during the course of a lesson through further questioning and guidance. The better the quality of the task, the better the quality of the interventions.

Differences between learning tasks can be understood in terms of a spectrum. At one end are "closed" tasks with a single well-defined outcome; at the other are "open" tasks with a wide range of acceptable outcomes. Tasks in language arts -- for example, the writing of a poem -- are mainly at the open end. But there are closed components even for such tasks -- for example, the observance of grammatical or genre conventions. Tasks in, say, mathematics are more often closed, but applications of mathematics to everyday problems can require open-ended evaluations. Thus, in varying measure, the guidance needed for these two types of learning work will be needed in all subjects.

Despite these differences, experience has shown that the

Working Inside the Black Box: Assessment for Learning in the Classroom.

generic skills that have been developed do apply across subjects. One of the project's science teachers gave a talk to the whole staff about his experiences using some of the generic skills that we've been discussing and subsequently found how such practices distributed themselves throughout the disciplines:

Art and drama teachers do it all the time, so do technology teachers (something to do with open-ended activities, long project times, and perhaps a less cramped curriculum?). But an English teacher came up to me today and said, "Yesterday afternoon was fantastic. I tried it today with my year 8s, and it works. No hands up, and giving them time to think. I had fantastic responses from kids who have barely spoken in class all year. They all wanted to say something, and the quality of answers was brilliant. This is the first time for ages that I've learnt something new that's going to make a real difference to my teaching." -- James, Two Bishops School

Motivation and Self-Esteem

Learning is not just a cognitive exercise: it involves the whole person. The need to motivate students is evident, but it is often assumed that offering such extrinsic rewards as grades, gold stars, and prizes is the best way to do it. However, there is ample evidence to challenge this assumption.

Students will invest effort in a task only 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, and those who have a track record as losers will see little point in trying. Thus the problem is to motivate everyone, even though 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. Here are a few examples:

* Students who are told that feedback "will help you to learn" learn more than those who are told that "how you do tells us how smart you are and what grades you'll get." The difference is greatest for low achievers.¹¹

* Students 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 outperforms the former.¹²

* In a competitive system, low achievers attribute their performance to lack of "ability"; high achievers, to their effort. In a task-oriented system, all attribute performance to effort, and learning is improved, particularly among low

achievers.¹³

* A comprehensive review of research studies of feedback found that feedback improved performance in 60% of the studies. In the cases where feedback was not helpful, the feedback turned out to be merely a judgment or grade with no indication of how to improve.¹⁴

In general, feedback given as rewards or grades enhances ego involvement rather than task involvement. It can focus students' attention on their "ability" rather than on the importance of effort, thus damaging the self-esteem of low achievers and leading to problems of "learned helplessness."¹⁵ 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.¹⁶

The Big Idea: Focus on Learning

Our experiences in the project all point to the need to rethink a teacher's core aim: enhancing student learning. To achieve this goal calls for a willingness to rethink the planning of lessons, together with a readiness to change the roles that both teacher and students play in supporting the learning process.

A Learning Environment: Principles and Plans

Improvement in classroom learning requires careful forethought.

Actually thinking about teaching has meant that I have been able to come up with ideas and strategies to cope with whatever has arisen and has contributed greatly to my professional development. I now think more about the content of the lesson. The influence has shifted from "What am I going to teach and what are the pupils going to do?" toward "How am I going to teach this and what are the pupils going to learn?" -- Susan, Waterford School

One purpose of a teacher's forethought is to plan to improve teaching actions. So, for example, the planning of questions and activities has to be in terms of their learning function.

I certainly did not spend sufficient time developing questions prior to commencing my formative training. . . . Not until you analyze your own questioning do you realize how poor it can be. I found myself using questions to fill time and asking questions which required little thought from the students. When talking to students, particularly those who are experiencing difficulties, it is important to ask questions which get them thinking about the topic and

Working Inside the Black Box: Assessment for Learning in the Classroom.

will allow them to make the next step in the learning process. -- Derek, Century Island School

Of equal importance is concern for the quality of the responses that teachers make, whether in dialogue or in feedback on written assignments. Effective feedback should make more explicit to students what is involved in a high-quality piece of work and what steps they need to take to improve. At the same time, feedback can enhance students' skills and strategies for effective learning.

There is also a deeper issue here. A learning environment has to be "engineered" to involve students more actively in the learning tasks. The emphasis has to be on students' thinking and making that thinking public. As one teacher put it:

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

Collaboration between teachers and students and between students and their peers can produce a supportive environment in which students can explore their own ideas, hear alternative ideas in the language of their peers, and evaluate them.

One technique has been to put the students into small groups and give each student a small part of the unit to explain to [his or her] colleagues. They are given a few minutes' preparation time, a few hints, and use of their exercise books. Then each student explains [his or her] chosen subject to the rest of the group. Students are quick to point out such things as, "I thought that the examples you chose were very good as they were not ones in our books. I don't think I would have thought of those." Or "I expected you to mention particles more when you were explaining the difference between liquids and gases." These sessions have proven invaluable -- not only to me, in being able to discover the level of understanding of some students, but to the students too. -- Philip, Century Island School

An additional advantage of such an environment is that a teacher can work intensively with one group, challenging the ideas and assumptions of its members, knowing that the rest of the class members are also working hard.

So the main actions to be taken to engineer an effective learning environment are:

- * Plan classroom activities to give students the opportunity to express their thinking so that feedback can help develop it;
- * formulate feedback so that it guides improvement in learning;
- * use activities that demand collaboration so that everyone is included and challenged and train students to listen to and respect one another's ideas; and
- * be sure that students are active participants in the lessons and emphasize 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.

A Learning Environment:

Roles and Expectations

It is one thing to plan new types of classroom activity and quite another to put them into practice in ways that are faithful to the aims they were developed to serve. Here there are no recipes to follow in a uniform way. Inside the Black Box was clear in stating that the effective development of formative assessment would come about only if "each teacher finds his or her own ways of incorporating the lessons and ideas that are set out above into her or his own patterns of classroom work."

A second principle is that 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 "classroom contract" between the teacher and the student -- the rules that govern the behaviors that are expected and seen as legitimate by teachers and students.

The students have to change from behaving as passive recipients of the knowledge offered by the teacher to becoming active learners who can take responsibility for and manage their own learning.

For the teachers, courage is necessary. One of the striking features of the project was that, in the early stages, many participants described the new approach as "scary" because they felt they were going to lose control of their classes. Toward the end of the project, they spoke not of losing control but of sharing responsibility for the students' learning with the class -- exactly the same process but viewed from two very different perspectives. In one

Working Inside the Black Box: Assessment for Learning in the Classroom.

perspective, the teachers and students 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

What has been happening here is that everybody's expectations -- that is, what teachers and students think that being a teacher or being a student requires you to do -- have been altered. While 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. For example, many teachers started by using questions to encourage thinking. Then they improved their oral and written feedback so that it brought thinking forward and went on to develop peer and self-assessment.

To summarize, expectations and classroom culture can be changed:

- * by changing the "classroom contract" so that all expect that teacher and students work together for the same end: the improvement of everyone's learning;

- * by empowering students to become active learners, thus taking responsibility for their own learning;

- * by incorporating the changes in the teacher's role one step at a time, as they seem appropriate; and

- * by sustained attention to and reflection on ways in which assessment can support learning.

What You Can Do

To incorporate some of the ideas about formative assessment into your own practice, the first step is to reflect on what you are now doing. Discussion with colleagues and observation of one another's lessons can help spark such reflection.

A next step must be to try out changes. Wholesale change can be too risky and demanding, so it is often best to think of one thing you feel confident to try -- be it "traffic lights," peer assessment, improved questioning, whatever -- and simply try it. If you are a teacher in a middle school or high school, try it with just one group. Or if you are an elementary teacher, try it in just one subject area. We found that, as teachers gained confidence in the power of

allowing students to say what they know and what they need to know, the teachers decided that they should extend assessment for learning to the whole of their teaching.

Taking on further strategies will then lead to further progress. When several colleagues are collaborating, each starts with a different strategy and then shares findings. This process should lead to the explicit formulation of an "action plan," comprising a range of strategies to be used, in combination, preferably starting with a class at the beginning of the school year. The first reason to start at the beginning of the year is so that there can be time to accustom both teacher and students to a new way of working. The second is that it can be very difficult to change the established habits and routines in the middle of a year. The experience of a year's sustained work, with only a few classes, preferably alongside similar efforts by colleagues, can provide a firm basis for subsequent adoption of new practices on a wider scale.

Collaboration with a group trying out similar innovations is almost essential. Mutual observation and the sharing of ideas and experiences about the progress of action plans can provide the necessary support both with the tactics and at a strategic level. Support for colleagues is particularly important in overcoming those initial uncertainties when engaging in the risky business of changing the culture and expectations in the classroom.

As for any innovation, support from administrators is essential. One way administrators can support change of this kind is to help peer groups of teachers find time to meet on a regular basis. Opportunities should also be found for teachers to report to faculty and staff meetings.

The work of any group experimenting with innovations is an investment for the whole school, so that support should not be treated as indulgence for idiosyncratic practices. Indeed, such work should be integrated into a school improvement plan, with the expectation that the dissemination of fruitful practices will follow from the evaluation of a group's experiences.

At the same time, there may be a need to review current school policies because such policies can actually constrain the use of formative assessment. A notable example would be a policy that, by demanding that a score or grade be given on every piece of homework, prevents the serious use of comments. Five of the six schools in the KMOFAP project have, following the experience of their science and mathematics teachers, modified their policies to allow "comment only" marking; for two of these, the modification was that no scores or grades be given on homework throughout the school. In another example, a

Working Inside the Black Box: Assessment for Learning in the Classroom.

"target setting" system that required very frequent review was inhibiting any change in learning methods that might slow down immediate "progress" in order to produce medium- to long-term gains in learning skills. Those engaged in innovations may need formal exemption from such policies.

Thus support, evaluation, and subsequent dissemination of innovation in assessment for learning will be planned in a coherent way only if the responsibility for strategic oversight of the development is assigned to a member of the school leadership team. Our experience supports the view that to realize the promise of formative assessment by leaving a few keen individuals to get on with it would be unfair to them, while to do it by imposing a policy that requires all teachers to immediately change their personal roles and styles would be absurd.

What is needed is a plan, extending over at least three years, in which a few small groups are supported for a two-year exploration. These groups then form a nucleus of experience and expertise for disseminating their ideas throughout the school and for supporting colleagues in making similar explorations for themselves.

1. Paul Black and Dylan Wiliam, "Inside the Black Box: Raising Standards Through Classroom Assessment," Phi Delta Kappan, October 1998, pp. 139- 48. A version of this article has been published and widely sold in the United Kingdom. A booklet, published in 2002, has also been widely distributed in the UK. It covers the same issues as the article and bears the same title with the same authors. Both booklets, and further booklets in this series, are published by NFER-NELSON.

Only a few references to the literature are given here. Further information about publications and other resources can be obtained on the King's College website in the pages of the King's Formative Assessment Group. Some of the publications can be downloaded from this site: www.Kel.ac.uk/education/research/Kal.html.

2. Paul Black and Dylan Wiliam, "Assessment and Classroom Learning," Assessment in Education, March 1998, pp. 7-71.

3. Paul Black, Christine Harrison, Clare Lee, Bethan Marshall, and Dylan Wiliam, Assessment for Learning: Putting It into Practice (Buckingham, U.K.: Open University Press, 2003).

4. Mary Budd Rowe, "Wait Time and Rewards as Instructional Variables, Their Influence on Language, Logic, and Fate Control," Journal of Research in Science Teaching, vol. 11, 1974, pp. 81-94.

5. Ruth Butler, "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, 1988, pp. 1-14.

6. Royce Sadler, "Formative Assessment and the Design of Instructional Systems," Instructional Science, vol. 18, 1989, pp. 119-44.

7. Royce Sadler, "Formative Assessment: Revisiting the Territory," Assessment in Education, vol. 5, 1998, pp. 77-84.

8. See, for example, Paul W. Foos, Joseph J. Mora, and Sharon Tkacz, "Student Study Techniques and the Generation Effect," Journal of Educational Psychology, vol. 86, 1994, pp. 567-76; and Alison King, "Facilitating Elaborative Learning Through Guided Student-Generated Questioning," Educational Psychologist, vol. 27, 1992, pp. 111-26.

9. See, for example, Mike Askew et al., Effective Teachers of Numeracy: Final Report (London: King's College London, School of Education, 1997). In this study, there was no correlation between the progress made by elementary school students in arithmetic and the highest level of mathematics studied by the teacher. Indeed, there was a nonsignificant negative correlation between the two. The students who made the most progress were taught by teachers without high levels of subject knowledge, but who emphasized the connections between mathematics concepts.

10. Sadler, "Formative Assessment and the Design of Instructional Systems."

11. Richard S. Newman and Mahna T. Schwager, "Students' Help Seeking During Problem Solving: Effects of Grade, Goal, and Prior Achievement," American Educational Research Journal, vol. 32, 1995, pp. 352-76.

12. Ruth Butler, "Task-Involving and Ego-Involving Properties of Evaluation: Effects of Different Feedback Conditions on Motivational Perceptions, Interest, and Performance," Journal of Educational Psychology, vol. 79, 1987, pp. 474-82.

13. Rhonda G. Craven, Herbert W. Marsh, and Raymond L. Debus, "Effects of Internally Focused Feedback on Enhancement of Academic Self-Concept," Journal of Educational Psychology, vol. 83, 1991, pp. 17-27.

14. Avraham N. Kluger and Angelo DeNisi, "The Effects of Feedback Interventions on Performance: A Historical

Working Inside the Black Box: Assessment for Learning in the Classroom.

Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory," *Psychological Bulletin*, vol. 119, 1996, pp. 254-84.

15. Carol S. Dweck, "Motivational Processes Affecting Learning," *American Psychologist* (Special Issue: Psychological Science and Education), vol. 41, 1986, pp. 1040-48.

16. Carol S. Dweck, *Self-Theories: Their Role in Motivation, Personality, and Development* (Philadelphia: Psychology Press, 2000).

PAUL BLACK is a professor emeritus, Department of Education and Professional Studies, King's College London, where CHRISTINE HARRISON is a lecturer in science education. CLARE LEE is a teacher advisor for the Warwickshire County Council. BETHAN MARSHALL is a lecturer in English education, Department of Education and Professional Studies, School of Education, King's College London. DYLAN WILIAM is director of the Learning and Teaching Research Center, Educational Testing Service, Princeton, N.J. They wish to thank the Medway and Oxfordshire local education authorities and the schools directly involved in the project described here, especially the teachers who shouldered the greatest burdens. The work on which this article is based was supported by grants from the Nuffield Foundation and from the U.S. National Science Foundation, as part of the Stanford University Classroom Assessment Project to Improve Teaching and Learning (CAPITAL). But the opinions expressed are those of the authors. (c) 2004, Dylan Wiliam.

Assessment for Learning

Assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting students' 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 that teachers and their students can use as feedback in assessing themselves and one another and in modifying 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.