

Formative Assessment in Mathematics

Part 3: The Learner's Role

Dylan Wiliam

Equals: Mathematics and Special Educational Needs 6(1) 19-22; Spring 2000

Introduction

This is the last of three articles reviewing the effectiveness of formative assessment, summarising the findings of a review of over 200 studies into formative assessment (Black and Wiliam, 1998). The first two articles dealt with the teacher's role in questioning and in giving feedback to learners. This last article focuses on the role of the *learner* in formative assessment, specifically the idea of sharing criteria with learners and student self-assessment. For each of these two ideas, I describe in detail below one experiment that has shown how effective involving students in these ways can be, and then go on to describe how they relate to research that we are currently doing with mathematics and science teachers.

Sharing criteria with learners

Frederiksen and White (1997) undertook a study of three teachers, each of whom taught 4 parallel Y8 classes in two US schools. The average size of the classes was 31. In order to assess the representativeness of the sample, all the students in the study were given a basic skills test, and their scores were close to the national average. All twelve classes followed a novel curriculum (called ThinkerTools) for a term. The curriculum had been designed to promote thinking in the science classroom through a focus on a series of seven scientific investigations (approximately two weeks each). Each investigation incorporated a series of evaluation activities. In half of each teacher's classes these evaluation episodes took the form of a discussion about what they liked and disliked about the topic. For the other two classes they engaged in a process of 'reflective assessment'. Through a series of small-group and individual activities, the students were introduced to the nine assessment criteria (each of which was assessed on a 5-point scale) that the teacher would use in evaluating their work. At the end of each episode within an investigation, the students were asked to assess their performance against two of the criteria, and at the end of the investigation, students had to assess their performance against all nine. Whenever they assessed themselves, they had to write a brief statement showing which aspects of their work formed the basis for their rating. At the end of each investigation, students presented their work to the class, and the students used the criteria to give each other feedback.

As well as the students' self-evaluations, the teachers also assessed each investigation, scoring both the quality of the presentation and the quality of the written report, each being scored on a 1 to 5 scale. The possible score on each of the seven investigations therefore ranged from 2 to 10.

The mean project scores achieved by the students in the two groups over the seven investigations are summarised in table 1, classified according to their score on the basic skills test.

Group	Score on basic skills test		
	Low	Intermediate	High
Likes and dislikes	4.6	5.9	6.6
Reflective assessment	6.7	7.2	7.4

Note: the 95% confidence interval for each of these means is approximately 0.5 either side of the mean

Table 1: Mean project scores for students

Two features are immediately apparent in these data. The first is that the mean scores are higher for the students doing 'reflective assessment', when compared with the control group—in other words, all students improved their scores when they thought about what it

was that was to count as good work. However, much more significantly, the difference between the 'likes and dislikes' group and the 'assessment' group was much greater for students with weak basic skills. This suggests that, at least in part, low achievement in schools is exacerbated by students' not understanding what it is they are meant to be doing—an interpretation borne out by the work of Eddie Gray and David Tall (1994), who have shown that 'low-attainers' often struggle because what they are trying to do is actually much harder than what the 'high-attainers' are doing. This study, and others like it, shows how important it is to ensure that students understand the criteria against which their work will be assessed. Otherwise we are in danger of producing students who do not understand what is important and what is not. As the old joke about project work has it: "four weeks on the cover and two on the contents".

Now although it is clear that students need to understand the standards against which their work will be assessed, the study by Frederiksen and White shows that the criteria themselves are only the starting point. At the beginning, the words do not have the meaning for the student that they have for the teacher. Just giving 'quality criteria' or 'success criteria' to students will not work, unless students have a chance to see what this might mean in the context of their own work.

Because we understand the meanings of the criteria that we work with, it is tempting to think of them as *definitions* of quality, but in truth, they are more like labels we use to talk about ideas in our heads. For example, 'being systematic' in an investigation is not something we can define explicitly, but we can help students develop what Guy Claxton calls a 'nose for quality'.

One of the easiest ways of doing this is to do what Frederiksen and White did. Marking schemes are shared with students, but they are given time to think through, in discussion with others, what this might mean in practice, applied to their own work. We shouldn't assume that the students will understand these right away, but the criteria will provide a focus for negotiating with students about what counts as quality in the mathematics classroom.

Another way of helping students understand the criteria for success is, before asking the students to embark on (say) an investigation, to get them to look at the work of other students (suitably anonymised) on similar (although not, of course the same) investigations. In small groups, they can then be asked to decide which of pieces of students' work are good investigations, and why. It is not necessary, or even desirable, for the students to come to firm conclusions and a definition of quality—what is crucial is that they have an opportunity to explore notions of 'quality' for themselves. Spending time looking at other students' work, rather than producing their own work, may seem like 'time off-task', but the evidence is that it is a considerable benefit, particularly for 'low-attainers'.

Student self-assessment

Whether students can really assess their own performance objectively is a matter of heated debate, but very often the debate takes place at cross-purposes. Opponents of self-assessment say that students cannot possibly assess their own performance objectively, but this is an argument about *summative* self-assessment, and no-one is seriously suggesting that students ought to be able to write their own GCSE certificates! Advocates of self-assessment point out that accuracy is a secondary concern—what really matters is whether self-assessment can enhance learning.

The power of student self-assessment is shown very clearly in an experiment by Fontana and Fernandez (1994). A group of 25 Portuguese primary school teachers met for two hours each week over a twenty-week period during which they were trained in the use of a structured approach to student self-assessment. The approach to self-assessment involved an exploratory component and a prescriptive component. In the exploratory component, each day, at a set time, students organised and carried out individual plans of work, choosing tasks from a range offered to them by the teacher, and had to evaluate their performance against their plans once each week. The progression within the exploratory component had two strands—over the twenty weeks, the tasks and areas in which the students worked were to

take on the student's own ideas more and more, and secondly, the criteria that the students used to assess themselves were to become more objective and precise.

The prescriptive component took the form of a series of activities, organised hierarchically, with the choice of activity made by the teacher on the basis of diagnostic assessments of the students. During the first two weeks, children chose from a set of carefully structured tasks, and were then asked to assess themselves. For the next four weeks, students constructed their own mathematical problems following the patterns of those used in weeks 1 and 2, and evaluated them as before, but were required to identify any problems they had, and whether they had sought appropriate help from the teacher.

Over the next four weeks, students were given further sets of learning objectives by the teacher, and again had to devise problems, but now, they were not given examples by the teacher. Finally, in the last ten weeks, students were allowed to set their own learning objectives, to construct relevant mathematical problems, to select appropriate apparatus, and to identify suitable self-assessments.

Another 20 teachers, matched in terms of age, qualifications, experience, using the same curriculum scheme, for the same amount of time, and doing the same amount of inservice training, acted as a control group. The 354 students being taught by the 25 teachers using self-assessment, and the 313 students being taught by the 20 teachers acting as a control group were each given the same mathematics test at the beginning of the project, and again at the end of the project. Over the course of the experiment, the marks of the students taught by the control-group teachers improved by 7.8 marks. The marks of the students taught by the teachers developing self-assessment improved by 15 marks—almost twice as big an improvement.

Now the details of the particular approach to self-assessment are not given in the paper, and are in any case not that important—Portuguese primary schools are, after all, very different from British ones. However this is just one of a huge range of studies, in different countries, and looking at students of different ages, that have found a similar pattern. Involving students in assessing their own learning improves that learning.

Putting formative assessment into practice

At the moment we are working with 24 teachers (12 science teachers and 12 mathematics teachers) in six schools to see how the ideas about effective formative assessment we have synthesized from the research literature can be incorporated into day-to-day classroom practice. As well as improving questioning, comment-only marking and the use of students' work to exemplify quality, the teachers are trying out a number of strategies related to student self-assessment.

For example, half the teachers are using 'traffic-lights' or 'smiley faces' to develop students' self-assessment skills. The teacher identifies a number of objectives for the lesson, which are made as clear as possible to the students at the beginning of the lesson. At the end of the lesson, students are asked to indicate their understanding of each objective by a coloured blob or a face. This provides useful feedback to the teacher at two levels. She can see if there are any parts of the lesson that it would be worth re-doing with the whole class, but also she will get feedback about which students would particularly benefit from individual support. However, the real benefit of such a system is that it forces the student to reflect on what she or he has been learning.

Level of understanding	traffic light	smiley face
good understanding	green	
not sure	yellow (amber)	
don't understand at all	red	

This feature of 'mindfulness' is one of the crucial features of effective formative assessment—effective learning involves having most of the students thinking most of the

time. Effective questioning is that which engages all students in thinking, rather than remembering, and doesn't allow students to relax simply because they've just answered a question, which means that it can't be their turn again until everyone else has been asked a question.

This notion of 'mindfulness' also gives some clues about what sort of marking is most helpful. Many teachers say that formative feedback is less useful in mathematics, because an answer is either wrong or right. But even where answers are wrong or right, we can still encourage students to think. For example, rather than marking answers right and wrong and telling the students to do corrections, teachers could, instead, feed back saying simply "Three of these ten questions are wrong. Find out which ones and correct them". After all, we are often telling our students to check their work, but rarely help them develop the skills to do so.

Other teachers are experimenting with 'end-of-lesson' reviews. The idea here is that at the beginning of the lesson, one student is appointed as a 'rapporteur' for the lesson. The teacher then teaches a whole-class lesson on some topic, and finishes the lesson ten or fifteen minutes before the end of the lesson. The student rapporteur then gives a summary of the main points of the lesson, and tries to answer any remaining questions that students in the class may have. If he or she can't answer the questions, then the rapporteur asks members of the class to help out. What is surprising is that teachers who have tried this out have found that students are queuing up to play the role of rapporteur, provided this is started at the beginning of the school year, or even better, when students are new to the school (year 1, year 3 or year 7).

Summary

Although at first sight quite different, the four elements of effective formative assessment outlined in this and the previous two papers form a coherent set of strategies for raising achievement, particularly for low-attainers. Rich questioning and effective feedback focus on the teacher's role—first being clear about where we want students to get to, asking appropriate questions to find out where they are, and feeding back to students in ways that the students can use in improving their own performance. Sharing criteria with learners and student self-assessment focus on the learners role—first being clear about where they want to get to, and then monitoring their own progress towards that goal.

To be effective, these strategies must be embedded into the day-to-day life of the classroom, and must be integrated into whatever curriculum scheme is being used. That is why there can be no recipe that will work for everyone. Each teacher will have to find a way of incorporating these ideas into their own practice, and effective formative assessment will look very different in different classrooms. It will, however, have some distinguishing features. Students will be thinking more often than they are trying to remember something, they will believe that by working hard, they get cleverer, they will understand what they are working towards, and will know how they are progressing.

In some ways, this is an old-fashioned message—very similar to the 'good practice' guidelines that were published by HMI in the 1970s and 1980s. What is new is that we now have hard empirical evidence that quality learning does lead to higher achievement. Teachers do not have to choose between teaching well on the one hand and getting good results on the other. Even if all a school cares about is improving its national test scores and exam results, the evidence is that working on formative assessment is the best way to do it. The bonus is that it also leads to better quality learning.

References

Black, P. J. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice*, 5(1), 7-73.

Fontana, D. & Fernandes, M. (1994). Improvements in mathematics performance as a consequence of self-assessment in Portuguese primary school pupils. *British Journal of Educational Psychology*, 64, 407-417.

Frederiksen, J. R. & White, B. Y. (1997). Reflective assessment of students' research within an inquiry-based middle school science curriculum. Paper presented at the *Annual meeting of the American Educational Research Association*. Chicago, IL.

Gray, E. M. & Tall, D. O. (1994). Duality, ambiguity and flexibility: a 'proceptual' view of simple arithmetic. *Journal for Research in Mathematics Education*, **25**(2), 116-140.