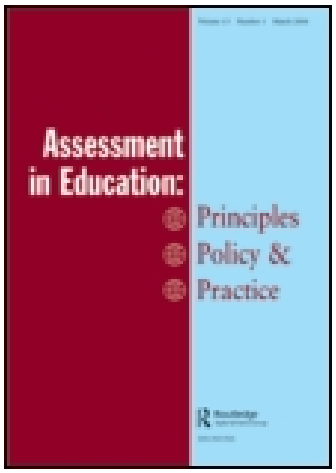

Assessment in Education: Principles, Policy & Practice
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/caie20

Formative Assessment: revisiting the territory

D. Royce Sadler [a]
[a] Griffith University, School of Education , Nathan, Queensland 4111, Australia
Published online: 28 Jul 2006.

To cite this article: D. Royce Sadler (1998) Formative Assessment: revisiting the territory, Assessment in Education: Principles, Policy & Practice, 5:1, 77-84, DOI: 10.1080/0969595980050104

To link to this article: http://dx.doi.org/10.1080/0969595980050104

Formative Assessment: revisiting the territory

D. ROYCE SADLER
Faculty of Education, Griffith University, Nathan, Queensland 4111, Australia

The review article by Black & Wiliam (1998, this issue) draws together research carried out since 1988 on the effectiveness of what is increasingly being termed formative assessment. This refers to assessment that is specifically intended to provide feedback on performance to improve and accelerate learning. The authors have a particular interest in empirical studies that are ecologically valid. Such studies reflect as closely as possible real teaching situations rather than being experiments based on special interventions that test a narrow element of a learning theory. Their welcome review paper covers research spanning all sectors of education from kindergarten through university, and follows two earlier reviews by Natriello (1987) and Crooks (1988). The focus on formative assessment necessarily meant that teacher-made assessments and self and peer assessment provided the principal criterion for selection of the studies reviewed.

Black and Wiliam found that, by and large, formative assessment is effective in virtually all educational settings: content areas, knowledge and skill types, and levels of education. The research also indicated that grades and marks do not deliver as much formative effectiveness as tailored comments, and in some situations can be counterproductive, particularly with learners of lower ability. What also emerged, however, was that the quality of feedback is a crucial issue. Not only is it of fundamental importance in understanding the role of formative assessment in improving learning, it is also often inadequately conceptualised and documented in research studies. Rigour in investigative design and analysis amounts to little if the nature of the treatment is poorly described. Hence directions for future development include both further analytic work and comprehensive empirical research.

One of the difficulties with experiments and quasi-experiments on the effectiveness of feedback on student learning is that the results may be delayed or masked by other factors. In particular, temporal conditioning of the students, that is, the long-term exposure of students to defective patterns of formative assessment and the socialisation of students into having to accept a wide variety of practices and teacher dispositions (many of which may appear incoherent or inconsistent), promote accommodating survival habits among students. To the extent that these are learned and embedded coping responses, they will take ingenuity, patience and time on the part of educators to reverse. Substantial modification to the learning environment through changes to regular classroom practice involves turning the learning culture around. Transformations made within one part of the curriculum (for example,

0969-594X/98/010077-08

feedback enhancement in one subject area with a particular teacher) may not be accompanied by complementary, mutually reinforcing teacher behaviours in other parts of the environment where other teachers follow a low feedback route. Any movement towards feedback-enhanced learning conditions (and investigations into their effectiveness, especially those based on experiments and quasi-experiments) must be carried out for long enough for the new procedures to be viewed by learners as normal and natural.

Students should also be trained in how to interpret feedback, how to make connections between the feedback and the characteristics of the work they produce, and how they can improve their work in the future. It cannot simply be assumed that when students are 'given feedback' they will know what to do with it.

In responding to Black & Wiliam's review, I comment on several themes that recur throughout their paper. The themes are not necessarily identified as such (certainly, they have no specific treatment as sections of their paper) but are, to some degree, those that presented themselves to me as they resonated with my own work and perspectives (Sadler, 1989). Contrary to what might be expected after several decades of research, there remains much that is unresolved and problematic, and much still to be done. It is possible, in fact, to trace a certain historical movement in thinking and experiments on formative assessment. Perhaps the first phase flowed from stimulus-response theory and identified feedback with knowledge of results. This could then lead to remediation (when incorrect) or reinforcement (when correct), higher motivation, and hence higher achievement. The second phase took a different tack: feedback was concerned with praise for effort, which would lead to higher self-esteem, more effort, and finally higher achievement. Recent research pays more attention to specific feedback tailored to both the nature of the assessment task and the learner's response to that task, progressive appreciation by the learner of what constitutes high quality work and the strategies needed to attain high standards, and thence high achievement.

Black & Wiliam recognise that the essential and necessary role of the teacher is to act as a mediator between, on the one hand, a body of knowledge and skills to be learned, and on the other hand, the learner. The knowledge base is inanimate, and in some instances is not rigidly fixed but still malleable. Often, how it is to be perceived and valued by teacher and learner is negotiable. The learner is, however, a person, a sentient being, situated in a context largely constructed by others. The role of the teacher could broadly be described as working to reduce (but not necessarily eliminate) the rate of error production in trial and error learning, and thereby to make learning more efficient. Teachers can do this effectively only if they know thoroughly both sides of the operation, and how to build bridges between the two. This view of teaching applies when the teacher is both responsible for, and accountable in the first instance to, the learner, and accepts that responsibility. As Black & Wiliam put it: 'All [classroom] work involves some degree of feedback between those taught and the teacher, and this is entailed in the quality of their interaction which is at the heart of pedagogy. The nature of these interactions between teachers and students, and of students with one another, will be key determinants for the outcomes of any changes...' (this issue, p. 16).

Of course, 'teaching' and 'learning' take place in a myriad of other contexts not connected with formal education, such as through instruction manuals, public lectures, broadcasting, advertising, and religious sermons.

Black & Wiliam's concern with ecological validity is important because, ultimately, we look for procedures that can be built routinely into learning contexts situated within schools and universities generally. It does not necessarily follow, however, that all of the research should be carried out by classroom teachers. Theoretically or practically desirable practices need to be informed by an adequate conceptualisation of what is supposed to go on. On the other hand, it would be dangerous to over-theorise the process and build elaborate schemas that ignore the contributions of competent and highly pragmatic practitioners.

In any case, whether feedback works (that is, leads to improved learning) is not the sole criterion for judging its desirability. Costs also come into the picture. Perhaps the primary issues are: How much of what we know ought to be done can actually be done with the resources available? How can the learning culture be turned gradually around to reflect better what we know (and will in the future know) about how to make feedback work for students? Costs are not, of course, solely monetary. They must also be appraised in terms of costs against the student: learning foregone as the impact of improper, inadequate or ineffective feedback is passed on and accumulated while students move through successive educational programmes. Even changes from one teacher to another have the potential for interrupting growth as momentum is lost. This loss may well be most serious for less able learners.

What is to be made of the apparently negative results from the research reviewed by Black & Wiliam that show either no improvement or, in the worst cases, a decrement in performance when the learning environment specifically incorporates feedback? If we adopt an optimistic position—which is to ask: How can good formative feedback *ever* backfire in the learning process and inhibit learning—the response must be not to question whether feedback should be provided, but how it should be provided, that is, its nature, contextualisation, timing and so on. Incorporating feedback is surely as fundamental a characteristic of responsible and responsive learning systems as having a teacher at all. This is not merely optimism, but a fundamental tenet in our understanding of what it means to teach.

At this point, I want to return to the quote from Black & Wiliam above, about good feedback lying at the heart of good pedagogy. The source of the feedback to facilitate learning is less important than its validity. (The source could be the teacher, or it could be a peer. Trust and personal interaction are always important elements.) Let me try to analyse this relationship in terms of both the process of evaluation, and what a highly competent teacher brings to this process.

The Feedback Ideal as a Model for Practice

What does a good teacher do in providing feedback to a student? This is an important question because any formative assessment that is not self-assessment involves communication. Assume for the present that the communication is (initially

at least) from the teacher to the learner. This is the most common situation. The communication is clearly not between equals, so the nature of the inequality needs to be recognised as part of an understanding of what makes for effective communication. We need to look both at what the teacher does to provide feedback to a student, and at what the teacher brings to assessment episodes to make that activity possible.

Three elements make up a typical teacher feedback act. First, the teacher must attend to the learner's production. Second, the teacher appraises this against some background, or reference framework. This process is invariably comparative, although sometimes what the production is being compared with is elusive. (On this point, I part company with Black & Wiliam, who assert that it is possible to evaluate something 'in its own terms' (this issue, p. 53). This may indeed seem to occur, but it is usually because the reference 'point' is not concrete but exists in some unarticulated or non-exemplified state inside the head of the assessor. I return to this issue in the next section.)

A teacher's evaluative comparison also usually involves some reflection and identification of strengths or weaknesses (as distinct from simply liking or disliking the production). Finally, to reflect the judgment, the teacher makes an explicit response, such as assigning the learner's work to a class (as in grading), mapping it on to a number line (marking), or providing a verbal statement about the quality itself (the reasons for the judgment and ways in which some of the shortcomings could be remedied).

Having identified the basic elements of an evaluative act, what do highly competent teachers bring to this act? What intellectual and experiential resources do they depend on? I list here six resources.

First, teachers bring superior knowledge about the content or substance of what is to be learned. Teachers commonly have a far more extensive and elaborate knowledge base than their students. This includes straightforward factual matters (for example, the author of a particular book, or the planet nearest the sun, or the common form of the normal equations in simple linear regression) that enable them to recognise immediately whether a particular student's response is correct, partially correct, or incorrect, or whether the idea of correctness makes any sense in the context. It also includes procedural knowledge (for example, the variety of ways to do something, and which ones are better than others) and what might be termed a connoisseur's knowledge of a field or discipline.

Second, teachers bring a set of attitudes or dispositions towards teaching as an activity, and towards learners, including their own ability to empathise with students who are learning, their desire to help students develop, improve and do better, their personal concern with the validity of feedback and the veracity of their own judgments, and their patterns in offering help (for example, a disposition to provide less feedback for works of brilliance, to provide considerable help for clearly salvageable cases, and to provide little help for apparently hopeless cases).

Third, teachers bring skill in constructing or compiling tests, in devising tasks, and generally in working out ways to elicit revealing and pertinent responses from students. Selecting from existing tasks or creating new ones demands refined

subjective judgments from the teacher. Such tasks must be sufficiently dissimilar from those previously attempted as learning exercises to test real achievement rather than memory and regurgitation (unless that is the legitimate aim). But they must also be similar enough to fall within the region that reasonably allows transfer or extended application of learning. The idea of teacher-as-assessor assumes that teachers know the learners and are themselves in a special sort of mastery position over the domain of knowledge. They therefore bring this as a resource to the teaching and assessing environment. By contrast, students mostly respond to already constructed tasks.

Having said that, it should not be assumed that this unidirectional practice of teacher-as-assessor is itself always justified or best. A strong case can be made that students should be taught how to change their pattern of thinking so that they know not only how to respond to and solve (externally sourced) problems but also how to frame problems themselves. They need this partly to guide their learning in between, or to prepare for, teacher assessments, but equally as part of their progressive journey into self assessment, and at more advanced levels, as a key skill for professional life. Indeed, Black & Wiliam cite research showing the efficacy of students being taught to construct their own tasks as part of the pedagogical environment.

Fourth, teachers bring a deep knowledge of criteria and standards appropriate to the assessment task. These criteria and standards may exist in an unarticulated form (which makes them difficult to share with learners) or in a more fully developed standards-referenced form (see Sadler, 1987). In addition, or often in place of, a clear idea of criteria and standards, teachers bring a set of expectations about what students should be able to produce in response to their (the teachers') assessment tasks. These expectations are a function of two elements. The first is a set of generalised expectations about the performance of the cohort of students being taught, based on a judgment of their ability levels. The second is the set of specific expectations about how the current cohort of students will respond to the immediate task, given recent teaching in the area and the students' experience with different and possibly less demanding assessment tasks.

Fifth, teachers bring evaluative skill or expertise in having made judgments about student efforts on similar tasks in the past. In normal teaching situations, teachers routinely make hundreds of qualitative judgments each year. This provides them with extensive, first-hand, current experience as assessors. The requirement that teachers make judgments demands that teachers attend conscientiously to the features of student performances. In non-convergent learning contexts, this automatically exposes teachers to a wide variety of ways in which the students approach problem solving, and how they argue, evaluate, create, analyse and synthesise. Many of these ways will be beyond the ability of the teacher to imagine, so that the teacher learns from the students. In turn, this exposure to other people's imaginations and strategies extends and enriches the teacher's repertoire of tactical moves. (This is especially important in contexts where teachers themselves are not in the habit of producing works of the same type as they require of students, for example, writing poetry, narratives, or investigative reports. The students' responses then provide vicarious experience about the challenges of production.) As all of this becomes part

of the teacher's tacit professional knowledge, it can be drawn upon on demand to provide helpful feedback to students.

Finally, teachers bring expertise in framing feedback statements for students. The simplest case, of course, is when a teacher tells a learner that something is correct or incorrect. That form of feedback can, in principle, be automated through objective testing and a key to responses, or more demandingly, through using an indexed body of material or look-up routines that make self-scoring possible for the learner. In more complex situations (even with children in lower grades), more sophisticated forms of feedback are called for. These may be of various types, but commonly take the form of written or oral statements that are intended to be interpreted by (and hence be cognitively accessible to) students. They include (non-evaluative) descriptions of the features of a student's work, evaluative comments linked to criteria that indicate those features that add to or detract from high quality, suggestions for alternative paths or arrangements that would have led to improvement, and exemplifying feedback that demonstrates (not just tells the student) in concrete terms an improved possible approach. In tailoring feedback for particular students, the teacher also draws on a knowledge of those persons' previous performances as well as their personalities.

An appreciation of these resources is important for two reasons. First, it is against this formidable array of the teacher's personal resources that studies on the effectiveness of feedback have to be seen. Ultimately, the intention of most educational systems is to help students not only grow in knowledge and expertise, but also to become progressively independent of the teacher for lifelong learning. Hence if teacher-supplied feedback is to give way to self assessment and self monitoring, some of what the teacher brings to the assessment act must itself become part of the curriculum for the student, not an accidental or inconsequential adjunct to it. The research reviewed by Black & Wiliam shows clearly that self and peer assessment hold great potential. They may become even more effective if students are specifically inducted into the processes of making sound qualitative judgments, and defending them. In other words, the processes and resources that are accepted as natural and normal for the professional teacher need to be replicated for the students and built into their learning environment. Obviously, teachers need to have professional preservice and inservice training for these specific requirements of formative assessment.

The second reason is that the teacher, by virtue of being a professional in a formal educational setting, is in a rich and authoritative position with respect to assessment. This position carries with it obligations to take into account the relatively poor position of the learners. Students' knowledge of the subject being learned is by definition partial. Hence any feedback must be expressed by the teacher in language that is already known and understood by the learner. It cannot extend beyond that to any significant degree. The attitude of the teacher towards helping students is also crucial, but is to a large extent beyond the control of the learner. And any tendency on the part of the teacher to provide differential levels of feedback for learners of different levels of performance (especially those at the lower end) treats students inequitably.

Devising assessment tasks represents in many instances creative and integrative activity of a high order. Learners see only the end result (the task to be completed), not the creative work, with its attempts and revisions, that go into devising the assessment task before it is administered. Learners often have little on which to base expectations about what should be delivered. Furthermore, their primary focus is on their own learning, and they frequently have little access to the performances of others, historic or current. Teachers, on the other hand, can and often do make adjustments to their expectations about how students should perform at a task after students have made their attempts. These adjustments are often made on the run, more or less intuitively, sometimes to correct for deficiencies in assessment task specifications, but mostly for the putative 'benefit' of the learners. Learners may have some experience at making judgments, particularly in contexts that promote self and peer assessment, but the number of those judgments is limited. Finally, students (again except for self-assessment and peer-assessment contexts) have limited opportunity to develop expertise in constructing evaluative statements, and thus to consolidate and clarify their own judgments. Given the differences between teachers' and learners' knowledge and experience in assessment, communication across the divide for formative purposes is an issue worthy of serious study.

The Question of Standards

For feedback to function effectively, effort must be put into creating learning environments where teachers' judgments are minimally contaminated by the performances of other (that is, current) students. It is impossible to make judgments about the quality of something *purely* 'in its own terms', that is, in the complete absence of any reference points or framework at all. But it is possible to define standards and to make consistent judgments against those standards. The critical issue is not whether particular forms of communication (such as grades) are used, because numbers and other symbols have norm-relevant implications only if we make them so.

Black & Wiliam explicitly link 'giving grades' with 'normative feedback' (this issue, p. 13), but this is not an inevitable connection. The issue is this: Can we develop a standards framework that can be implemented in ordinary classrooms which, at the point of application at least, is independent of the performance of other group members? Only if this route is followed can some of the reported negative effects of particular forms of feedback be avoided. If we ignore for the present the classical definition of feedback in relation to system control (where it is related to actions that lead to some closing of a gap between an actual and a stable reference state) and focus on the common interpretation in education (as critical, and hence negative, comment on a learner's performance), the reason for negative effects of negative feedback can be traced to inadequately specified standards.

To the extent that a teacher tries to work without clearly defined standards, and defaults to an existentially determined baseline derived from how other students perform, the teacher is unable to provide task-related, standards-oriented feedback. When a learner realises that the feedback is at least partly cohort-dependent,

negative comment can be discouraging and interpreted as personal criticism. A typical response has been to encourage teachers to praise what is good (indeed, to be sure to find characteristics to praise even if the overall production is weak). As the theory goes, this praise is then supposed to promote a positive self-image, which in turn produces high motivation, and hence high achievement and perhaps even a love of learning. But in practice the fallback position of praising the student (as a person) for effort, for trying, leads to praise for work of a quality that does not deserve it.

An alternative option, which does come through in the research cited by Black & Wiliam, is to dissociate ego-involving and task-involving feedback, that is, to push towards measuring true accomplishment, which in its own good time leads to improved self-esteem. A precondition for this to work, however, is the implementation of a system of assessment in which absolute improvement, when it is made, is clearly recognised. If progress is always measured, say, against an implicit cohort median, a particular student may actually be making improvement but have it masked, because the reference framework inexorably advances (upward) with time. The framework has to be separated from cohort performance and remain stable so that real (or absolute) improvement can be plotted for each student.

Conclusion

I conclude by reiterating the position reached by Black & Wiliam's review: Formative assessment *does* make a difference, and it is the quality, not just the quantity, of feedback that merits our closest attention. By quality of feedback, we now realise we have to understand not just the technical structure of the feedback (such as its accuracy, comprehensiveness and appropriateness) but also its accessibility to the learner (as a communication), its catalytic and coaching value, and its ability to inspire confidence and hope.

References

BLACK, P. & WILIAM, D. (1998) Assessment and classroom learning, *Assessment in Education*, 5, pp. 7–74, this issue.

CROOKS, T.J. (1988) The impact of classroom evaluation practices on students, *Review of Educational Research*, 58, pp. 438–481.

NATRIELLO, G. (1987) The impact of evaluation processes on students, *Educational Psychologist*, 22, pp. 155–175.

SADLER, D.R. (1987) Specifying and promulgating achievement standards, *Oxford Review of Education*, 13, pp. 191–209.

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