Editorial

From Frequent Testing to Interactive Dialogue: Diversity and Challenge

Paul Black
King's College, London, United Kingdom

The articles assembled in this special issue reflect very faithfully the diversity in what is now a rich field of development. But because they do this, they also highlight a problem: what it is about this topic that leads to the contributors describing such a diverse range of approaches to the practices of formative assessment?

One insight into the diversity can be gained from examining the contrast between the papers by (say) Glenn and by Brookhart, Moss, and Long. The former is concerned with a data-driven approach to improvement of learning, the latter with ways in which teachers interact in classrooms to build up their students' powers as independent learners. All use feedback from evidence to improve learning, but the variety of ways in which this can be done can reflect diversity, both in educators' educational values and in their beliefs about the link between instruction and learning.

Johnson and Green describe an assessment approach which is focussed on the formative use of tests, and the use of feedback from them to improve the teachers' instruction. They carefully detail the way teachers use the evidence with emphasis on modifying their own instruction. The accounts both Bonner and of Ponte and colleagues have a similar focus on frequent testing, but the first shows that within this practice there can be a variety of methods and levels at which feedback can be given to pupils, whilst in the second the thorough analysis does show that more informative feedback produces better results.

In studies of this type, there is often no clear distinction between the summative and the formative uses of test outcomes. The study reported by Brown and Harris highlights the potential dangers: a set of tools designed to improve learning through instruction was helpful when used to enrich the informal interactions between teachers and pupils, but when taken up by school managements as instruments to serve accountability, teachers came to see them as oppressive and in conflict with their formative practices.

So whilst data can clearly be helpful, they have to be designed to be diagnostic. An interesting feature of Glenn's inquiry is that the teachers both learnt from test data, in that it did not confirm a hypothesis about a cause of weak performance, yet realised that it gave them little help in developing ways to improve. One problem that is clear here is the choice of a way to take the step, from an analysis of the data, to the design of innovations in instruction to deal with the difficulties it reveals. This step requires the application of theory, and of priority of purpose, in relation to students' learning. There is also a deeper difficulty in that the most effective use of assessment for learning calls for use of feedback in ways that radically challenges common instruction practices.

Paul Black

An observer encountering the topic for the first time through these papers might well conclude that the various authors do not share any agreed conception of the role of formative assessment in classroom instruction. There is not a radical conflict here, but rather a spectrum of conceptions. At one end is the 'testing' approach, the use of summative tests to provide guidance to teachers about the strengths and weakness of their work: at this extreme, the assessment events are infrequent, the involvement of the students may be slight, and the interpretation of the evidence problematic because the data is not rich. If one wants to use the data formatively to affect pupils' approach to their learning, the testing may become more frequent, but unless the tests are designed to explore learning diagnostically, the data may be of little use and even, in the case of multiple choice tests, possibly harmful. A particular problem here is that the stresses, on both teachers and students, associated with summative tests, affect any formative intentions, and can undermine their interpretation and use for formative purposes.

At the other end of this spectrum is the interactive dialogue approach, with a focus on the to-and-fro of involvement in discussion between teacher and students and between students themselves: each time a student makes a contribution, the teacher or another student will make a response which should be designed to help advance the thinking, of the respondent and/or of the group as a whole. In this case the assessments are very frequent, the involvement of the students will be strong, the evidence is rich in that the potential meaning of each fragment is explored and may be revised or developed—the interpretation is constructed in and through the dialogue.

Most implementations of assessment for learning will lie somewhere in the spectrum between these two. Any attempt to decide which is *right* might be pointless, but a careful analysis of the research evidence could show that some practices are more profitable than others. Yet the various practices do not have to be mutually exclusive, and indeed a teacher's particular portfolio of formative practices might include components from many parts of the spectrum. What might be profitable would be to explore some of the many differences between the two perspectives. A difficulty here is that reviews of the evidence about formative assessment cover a range of practices developed in a range of contexts, so whilst they can serve as encouragement to develop innovations, they cannot guarantee that any innovation will be valid in its particular context of implementation. In particular, great care must be taken in quoting such evidence in support of practices which are not supported by any of the evidence reviewed, a fault which is evident in the rhetoric of some government initiatives in both the United Kingdom and the United States.

Two different issues are raised in the paper by McGatha, Bush, and Rakes. The authors recognise that one limitation of their study is that they were not able to conduct systematic observation of implementation in classrooms: the experience of my colleagues and of others (Dillon, 1994) is that self-report by teachers, without supporting evidence, may be optimistically biased: feedback from classroom observations by researchers or colleagues is an important support and spur in achieving change.

This paper also adds to the findings of many other studies—that teachers need support from, and frequent interaction with, colleagues over an extended period of time if they are to make significant changes. To practice formative interaction through enriched dialogue can involve, for many teachers, a quite radical change in their practice, and even in their view of their role as a teacher. This is brought out in the accounts by teachers who have adopted this approach (Black et. al., 2003), and given their view of their purpose, it had to be taken seriously in McGatha, Bush, and Rakes' study. This involves attention to the detail of instruction: their finding that the detail of the

types of question asked is all-important is very significant here. The uneven pattern of learning gains is also significant. In part it may show that the differences between individual teachers are at comparable to any difference that this type of innovation may make.

The notion of a spectrum of practice, between the testing and the dialogue poles, is illustrated in a different way by Van Haneghan's study. This paper makes an important contribution, not only because it recognises how complex and potentially confusing the field has now become, but also because it proposes a way forward for representing the several dimensions of this complexity. Their report brings out that the underlying learning model on which instructional practice is based is clearly relevant for the validity of any assessments. It is, for example, important when a teacher has to decide, in the midst of a classroom discussion, how to interpret and so respond to a student's suggestion, particularly an unusual one: if the response is to be formative; that is, to help the student and others think more deeply, and not merely to pass judgment on the answer, then the teacher has to think about how the student may have come to make that suggestion, and frame a response that will either challenge the thinking, or will involve the student, or will enhance participation to involve others in a class; the aim being to explore and develop the reasoning behind the suggestion. This focus on what Van Haneghan calls the 'teachable moment' draws attention to the opportunities, but also the difficulties, in the conduct of formative dialogue.

This warning note is supported by the ambitious review reported by Ziebarth and colleagues, which is valuable in drawing attention both to the long and varied history of the development of assessment for learning, and to the strong investment, notably by the National Council of Teachers of Mathematics, in giving encouragement and support to teachers for its development. There are several important features, notably that teachers and students must be jointly involved, that it involves a range of classroom practices, and that what is new is that it has become a 'focal point of good pedagogy.' Given all of this, it seems disappointing that development of the concept in both higher education and in high schools seems so weak, and that it is given little attention in the programmes of teacher training which were surveyed. One notable weakness is the focus on the use of assessment to grade students rather than to involve them and help them to understand and to learn. They also correctly draw attention to the paucity of classroom studies: it is to be hoped that their plan to develop such studies, together with the other lines of inquiry which they have identified, can be put into effect.

Many classroom activities, and many forms of assessment, are not designed for a formative purpose—to serve it they have to be so designed that one can make valid inferences from the response about the thinking that lay behind that response. Even the most expensive method—an interactive and exploratory dialogue between teacher and pupils—is not without its problems because both participants have to continually engage in achieving understanding of what the others are saying. This issue is brought out in Van Haneghan's discussion of the work of Frederiksen and Collins.

This leads on to a very significant point made in Popham's book, as highlighted in Barney and McCowen's review, that teachers must move, from emphasis on competition in the classroom, to emphasis on the shared responsibility between, and the engagement of, students. In this respect, it is striking, and not comforting, that none of these papers refers to concerns that the testing-competition culture can have positively harmful effects on pupils' beliefs and attitudes concerning their own learning. The literature here deserves more attention (e.g., Dweck, 2000; Harlen & Deakin-Crick, 2003).

The paper by Brookhart, Moss, and Long is at the dialogue end of my spectrum. It stands out for its emphasis on the role of students in formative assessment. It is also distinctive in its report on their carefully designed programme of professional development. A key insight, significantly one expressed by a teacher about her own learning, features in one of their quotations:

> At first I thought this was an obvious point, but as I began thinking about my teaching I began to notice that although I stated the target, some of the students didn't internalize it to understand how the target pertained to them as a learner.

This paper is not about testing at all—it's about changing teachers' interactions with their pupils, with a recognition that a sustained programme is needed to produce any effect. The focus here has shifted—away from the teacher and towards the students. Their understanding of learning goals and their development of self-assessment is centre-stage. It is surely right that this should come before, and should be the benchmark for, any concerns with teachers practices.

The authors rightly state that their concern was with 'high impact formative practices', but they keenly appreciate that these are very challenging for teachers. Feedback on frequent testing is far less challenging, but its value in developing the learning autonomy of students has to be questioned.

References

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

Butler, R. (1988). Enhancing and undermining intrinsic motivation; the effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology, 58*(1), 1-14.

Dillon, J. T. (1994). *Using discussion in classrooms.* Buckingham: Open University Press.

Dweck, C. S. (2000). *Self-theories: their role in motivation, personality and development.* Philadelphia, PA: Psychology Press

Harlen, W. & Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education*, *10*(2), 169-208.