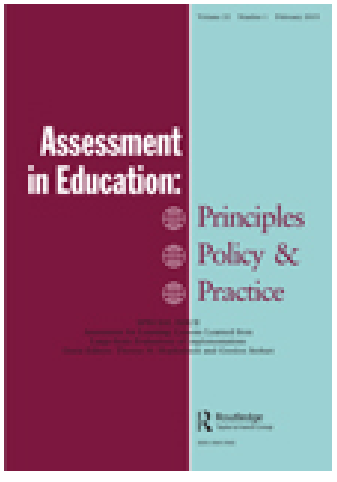

Assessment in Education: Principles, Policy & Practice

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/caie20

Formative assessment - an optimistic but incomplete vision

Paul Black[a]

[a] Department of Education and Professional Studies, King's College London, London, UK

Published online: 30 Jan 2015.

To cite this article: Paul Black (2015) Formative assessment - an optimistic but incomplete vision, Assessment in Education: Principles, Policy & Practice, 22:1, 161-177, DOI: 10.1080/0969594X.2014.999643

To link to this article: http://dx.doi.org/10.1080/0969594X.2014.999643

Assessment in Education: Principles, Policy & Practice, 2015
Vol. 22, No. 1, 161–177, http://dx.doi.org/10.1080/0969594X.2014.999643

Formative assessment – an optimistic but incomplete vision

Paul Black*

Department of Education and Professional Studies, King's College London, London, UK

(*Received 15 December 2014; accepted 15 December 2014*)

The preceding articles in this issue describe a diverse range of projects which had in common the aim of implementing or improving the practice of formative assessment, and thereby to secure some of the benefits attributed to it. This article attempts to set up a framework within which each of the different studies may be located and inter-related. There are three main sections. The first deals with the roles of assessment, both formative and summative, within a comprehensive model of pedagogy. The second considers the specific ways in which the different practices of assessment feedback help to develop the capacity of each student to become a thoughtful and independent learner. The third reviews the ways in which new assessment practices present problems to teachers in challenging them to re-think their role and similarly to students, when for both groups, new practices affect their ways of coping in the classroom.

Keywords: assessment for learning; summative assessment; theories of learning; learning to learn; assessment in pedagogy; teacher change; student change

Introduction

The 1998 review by Black and Wiliam held out a promise of ways to innovate teaching practices which had been shown to enhance students' learning. The eight articles in this special edition are evidence of that review's impact in their description of how that publication has influenced policy and practices in eight different countries. However, they are far from presenting a tale of overall success – the most optimistic claim amongst them is of 'partial success'. The review did qualify its message, stating on the one hand that 'there is enough evidence in place for giving helpful guidance to practical action' (p. 61), but on the other hand that:

> What does emerge is a set of guiding principles, with the general caveat that the changes in classroom practice that are needed are central rather than marginal, and have to be incorporated by each teacher into his or her practice in his or her own way. (Black & Wiliam, 1998a, p. 62)

In this article, I shall first reflect on this caveat in relation to a summary of the most important of the many problematic issues that the eight articles describe, thereby raising the question of whether these problems might have been foreseen or predicted. In the same section, I shall highlight a few components of the several implementations which have been described, drawing attention to some difficulties arising both from their diversity and because some important aspects are not reported.

*Email: paul.black@kcl.ac.uk

This discussion will lead into the next section which analyses, at a more general level, the roles of assessment in a comprehensive model of pedagogy, in terms of which I shall discuss the relationship between formative and summative assessment. It will be suggested that in order to tackle the problems of that relationship, an approach using marriage guidance rather than divorce must be chosen. The natural successor to this section will be a section analysing how assessments aid the development of students as effective and responsible learners. This will then lead to a section which will discuss the challenges which new approaches to assessment present to both teachers and students. These sections will constitute an attempt to map out a framework, which includes the main principles and practical determinants in terms of which any innovation report might be evaluated. The closing section will return to highlight some of the problems of designing projects of the type described in the eight papers, and to consider how lessons learnt from these may serve as useful guides for future work.

For the purpose of this article, I shall refer to the eight papers in this Special Issue by the names of their countries of origin.[1] I shall also include in my discussions the findings of the King's Medway Oxfordshire Formative Assessment Project (which I shall refer to below as the KMOFAP project), treating it, alongside the eight articles, as another attempt to implement the practices of assessment for learning. The findings of that project were published in four research articles (Black & Wiliam, 2003; Harrison, 2005; Lee & Wiliam, 2003; Wiliam, Lee, Harrison, & Black, 2004), and in a book for teachers (Black, Harrison, Lee, Marshall, & Wiliam, 2003). It involved a comparatively small group of 6 schools and about 40 teachers. I include discussion of its findings here, because I note that three of its five reports are not referred to in any of the articles in this Special Edition, and that no article refers to more than two of them.

The uneven spectrum of achievement

In all of the eight articles, the central role of assessment for learning appears in one form or another, although more explicitly in several. All have been motivated, by evidence, to promote assessment for learning, using in varying ways methods already explored elsewhere. In these respects, these articles add to similar studies published in the last 10 years: the range of these studies and the varieties of approaches that they have selected from previous work is carefully reviewed in the paper from Chile (Flórez Petour, 2015). The article from Canada adds to this with a review of the ways in which these attempts at implementation have 'gone wrong' (DeLuca, Klinger, Pyper, & Woods, 2015), while the account from Sweden draws attention to ambiguity in the definition of assessment for learning and to the fact that the procedures and instruments have not been precisely defined (Jonsson, Lundahl, & Holmgren, 2015).

The eight studies report on the diverse set of strategies for promoting innovation in assessment for learning, due in part to differences, both in the starting points and in the cultural contexts, between the eight countries. However, the information they supply is lacking in certain important features. For example, six of the papers do not give many details of the methods used by teachers to enhance participation of pupils in whole-class discussions – the USA study being a notable exception (Wylie & Lyon, 2015). Most of the reports rely on teachers' accounts of their classroom discussions, and whilst one mentions the use of classroom observation, none gives

direct evidence of classroom dialogue. Thus, the lessons that can be learned from these papers, about the detail of the changes actually achieved, are of limited scope.

Given the wide range and differing approaches to developing formative assessment practices, it is hardly surprising that many difficulties have been encountered. Indeed five of the articles show, at best, partial success whilst two others are about situations which are seriously beset with problems. Common to all is the tension between the investment in formative teaching and the pressures of the testing instruments used to satisfy demands for accountability. Another issue is the difference between top-down models of innovation and bottom-up initiatives; yet, the top-down approaches differ across a spectrum between supported exploration and imposed recipes. It is not surprising that, in the latter cases, teachers were less inclined to cooperate. On the other hand, positive evidence for the Canadian initiative was illustrated by one teacher saying 'I have completely changed my style of teaching' and another saying 'it's not just about sharing success criteria and learning goals, it's now about how we are teaching'.

A more general problem is the lack of understanding of the complexity and slow pace of teacher change. This was emphasised in the earliest commentary on the implications of the 1998 review article, published under the title *Inside the Black Box* (Black & Wiliam, 1998b, p. 15). However, it may be said that the 1998 review article was too optimistic where it said that there was enough evidence to justify applying the research findings to practical action. In an article reflecting on the lessons learnt from the KMOFAP project, Black and Wiliam commented on their 1998 articles that:

> … we inevitably, at some points, went beyond the evidence, relying on our experience of many years' work in the field. If we had restricted ourselves to only those policy implications that followed logically and inevitably from the research evidence, we would have been able to say very little. (Black & Wiliam, 2003, p. 628)

And later in the same article:

> Education research can and does make a difference, but it will succeed only if we recognize its messy, fragile, contingent nature. (p. 635)

Assessment in pedagogy

Many writers about assessment, and many teachers, regard assessment as a peripheral component of pedagogy, one that is inescapable but which always threatens to undermine the most valued aim, that of developing the learning capacity of their students. The phrase 'assessment for learning' challenges this view, and some handle this challenge by regarding it as quite separate from summative assessment. This I regard as a fundamental error, one that arises from the lack of a broad and more complex view of the role of assessment in pedagogy: one purpose of this section is to support this view.

In his commentary on the Black and Wiliam (1998a) review, Perrenound included the following statements about the focus on feedback in that article:

> This no longer seems to me, however, to be central to the issue. It would seem more important to concentrate on the theoretical models of learning and its regulation and their implementation. These constitute the real systems of thought and action, in which feedback is only one element. (Perrenoud, 1998, p. 86)

I did not take this statement seriously at the time, in part because, whilst it seemed then to be the only serious criticism of the review, it was not clear to me how the review could have been altered or supplemented to meet the criticism. However, the clue to that question could have been found in a later statement in Perrenoud's commentary:

> I would like to suggest several ways forward, based on distinguishing two levels of the management of situations which favour the interactive regulation of learning processes:
>
> - the first relates to the setting up of such situations through much larger mechanisms and classroom management.
> - the second relates to interactive regulation which takes place through didactic situations. (Perrenoud, 1998, p. 92)

A response to Perrenoud's challenges must include consideration of two aspects of his argument. The first should be an analysis of 'the mechanisms of classroom management', and the second should respond to the need to focus on 'the theoretical models of learning and its regulation'. I shall try to consider both aspects in this section and in the one following.

In many of the classic texts about theories of pedagogy, assessment seems to play a very small part. Arising out of consideration of this problem, and drawing on Hallam and Ireson (1999) and Wiske (1999), I proposed (Black, 2013) a simple model in terms of the following five stages:

(1) Clear aims: The first stage of planning.
(2) Planning activities: Setting up activities with the potential to achieve the aims.
(3) Implementation in the classroom.
(4) Review of the learning: Using informal assessments to check achievement.
(5) Summing up: Using assessment to guide decisions about the next stage of students' work.

This can be seen as a sequence in time of five main stages: however, there are also important interactions, in both directions, between the stages. The third stage is where interactive dialogue first operates, and presents challenges because students' responses are often unpredictable, and may be disruptive. However, the potential of any activity to engage the attention of students, in ways that can help steer a discussion towards the intended aims, has to be foreseen: such foresight is a key professional skill needed in stage 2. For the planning of this stage, the teacher has to combine a clear view of the relevant aims from stage 1 with experience of what may or may not work, with the particular class involved, in stage 3.

Stage 4 is the one which deserves more prominence than it is often given. A test near the end of the work on a topic should serve as a review, in part to check for problems encountered en route which are still unsolved: for this purpose, it should be near, not at, the end, in order to allow time for any work needed to deal with the problems it might reveal. However, there is also another purpose, in that such a review can help learners to achieve an overview through which they might see how different aspects inter-relate. Feedback on written work can be seen as a bridge between stages 3 and 4; one piece of written work may, in some cases, serve to review the work of several lessons, and conversely, collections of test scripts can be treated, through comment-only marking and/or peer-assessment, as occasions for

medium-term feedback. The main point is that in this stage, the summative assessment can serve, through the specific use of interactive dialogue, as a necessary stage in the learning. By contrast, if all that a test gives the students is marks or grades, then these functions are ignored.

One teacher in England, who responded very positively in the work of the KMOFAP project to develop formative assessment, found it difficult to maintain the positive approach to summative assessments in his subsequent work, as he explained several years later:

> With pressure 'from above', at all levels, most schools were content to 'tick' the AfL box rather than focus on developing and evolving formative practice. At times, I am certain, all of those who had been involved in the King's project felt like screaming.
>
> Sadly, my experience has been that AfL has been lost under and within the 'tickbox' culture that seems to have arisen in secondary schools, hand in hand with league tables. (Spenceley, 2009, p. 9)

Such problems arise because Stage 5 is different from the others, in that it has a particular function in guiding decisions, and in producing information to serve such guidance: those who can use this information may include students themselves, parents, those who may be teaching the student in the subsequent year, school leaders, and those who may choose or select the student for work or study beyond the school. However, this last stage can be problematic, and the effects of various levels of accountability pressures stand out clearly in the eight articles. In the Canadian exercise they seem to play only a minor part, whereas in Chile their dominance seems to prevent any serious development of formative practice; the study from Norway expresses the problem as follows:

> … the study has shown that there is a constant struggle involving teachers and policy makers regarding the need for trust in the system and the need for accountability.

In the KMOFAP project, the teachers concentrated on one of the school years in which there were no external accountability tests. However, when the KMOFAP leadership said, to the teachers involved in the project, that it would not include any work on summative tests, the teachers said that it would be unrealistic to attempt a complete separation of the summative from the formative, for they had to meet their own needs, and requirements set by the school, for producing summative results. This led to very productive work on the formative use of summative tests (Black et al., 2003, pp. 53–57) which concluded that:

> summative tests should be, and should be seen to be, a positive part of the learning process. (p. 56)

However, it was also evident that teachers lacked the skills to compose their own summative assessments, and lacked the confidence to compose their own, rather than use tests copied from other sources. A later attempt to tackle this problem in another two-year project did confront this problem by helping teachers to produce their own reliable, and more valid, assessments – but again this was only developed for school years when there were no accountability pressures (Black, Harrison, Hodgen, Marshall, & Serret, 2011).

The underlying problem here is the conflict between the responsibilities of teachers and schools for summative assessments of their students, and the responsibilities placed on them by external, accountability-loaded, tests to ensure the best possible

results for their pupils. In such situations, a lack of alignment between stages 4 and 5 is inevitable. This problem can in principle be resolved by placing responsibility for terminal and high-stakes assessments in the hands of teachers and schools. This situation has been mainly achieved in some states in Australia (Klenowski & Wyatt-Smith, 2014). The Scotland paper gives an account of how policy contradictions arose when the demand for data derived from national tests was dropped at national level, but sustained by district authorities (Hayward, 2015). Furthermore, new national policy called for alignment of curriculum and assessment, but failed for some time to give schools either advice or support about how to achieve such alignment with the successful implementation of assessment for education, which many had already achieved in Scotland's use, as its model strategy, of the KMOFAP project (Hallam, Kirton, Peffers, Robertson, & Stobart, 2004). What was needed here was help with the link between stages 1 and 2 and stages 3 and 4.

Instead, the extra burdens that the curriculum placed on schools were such that assessments eventually became the 'villains of work-load'.

In England, the long-established inclusion of school-based assessments in the main high-stakes assessments has recently been all-but abandoned. By contrast, the account of the situation in Trinidad and Tobago shows that their teachers have responsibility for summative assessments; however, they lack the confidence and the training to exercise that responsibility by any other means than imposing, in their school summative assessments, those features of externally based testing which are inimical to good teaching and learning (De Lisle, 2015). Thus, they routinely record marks but make little use of the data: assessment for learning is not consistent with their established beliefs and practices so that synergy between the formative and summative functions of assessment cannot be achieved.

This does not contradict lessons in this sphere which have been learnt in other countries, for their experience is that it takes several years of in-service training of teachers, and a system to support inter- and intra-school collaboration, if valid and trustworthy systems for summative assessment by teachers are to be firmly established. Even in these cases there has arisen, in recent years, a desire by some politicians to raise the scores of their country in the international 'league tables' which are the product of the work by TIMSS and PISA, a concern which exerts pressure on all to teach to these tests rather than to the targets set up in their own national systems.

These problems with Stage 5 are particularly unfortunate because it has a more general importance that is both obvious and yet rarely explored. This is the link between Stage 5 and the aims set out in Stage 1. Wyatt-Smith and Bridges (2008) described this link as follows:

> So basically once you have the assessment firmly in place the pedagogy becomes really clear because your pedagogy has to support that – that sort of quality assessment task … that was a bit of a shift from what's usually done, usually assessment is that thing that you attach on the end of the unit whereas as opposed to sort of being the driver which it has now become. (p. 48)

It is tempting to state the aims of a teaching episode in attractively general terms, but it is in the instruments used for the Stage 5 assessment that the meaning of these aims has to be made explicit. If the initial formulation of aims is not expressed with a clear link to what is assessed in Stage 5, these can be unhelpful, even misleading. Where teachers determine Stages 1 and 5, they have to work to achieve synergy

between them: if stage 5 is determined by external agencies, then teachers have to 'teach-to-the-test'.

Assessment and learning

The importance of the connection between assessment and learning was recognised in the KMOFAP project, and expressed as follows:

> At a more general level, however, it could be seen that the practical activities developed did implement the principles of learning that are prominent in the psychology literature. Examples are the constructivist principle, that learning action must start from the learner's existing knowledge, the need for active and responsible involvement of the learner, the need to establish in the classroom a community of subject discourse, and the value of developing meta-cognition. (Black & Wiliam, 2003, p. 634)

However, this very general statement was not followed up in more detail. In this section, some of the detail will be explored in the context of the five stages set out in the previous section.

Engaging in interactive dialogue

The first link is at stage 3, in which learners' capacity to engage in and learn from interactive dialogue should be developed. This link emerges clearly in several of the eight studies. The basic justification here was clearly expressed by Alexander as follows:

> Children, we now know, need to talk, and to experience a rich diet of spoken language, in order to think and to learn. Reading, writing and number may be acknowledged curriculum 'basics', but talk is arguably the true foundation of learning. (Alexander, 2006, p. 9)

This view he developed later in a more fundamental explanation:

> Talk vitally mediates the cognitive and cultural spaces between adult and child, among children themselves, between teacher and learner, between society and the individual, between what the child knows and understands and what he or she has yet to know and understand. (Alexander, 2008, p. 92)

In this view, involvement in dialogue both helps develop the understanding by students of the topic under discussion, and also helps develop their capacity to learn. This aspect is explained further by Wood:

> Vygotsky, as we have already seen, argues that such external and social activities are gradually internalized by the child as he comes to regulate his own internal activity. Such encounters are the source of experiences which eventually create the 'inner dialogues' that form the process of mental self-regulation. Viewed in this way, learning is taking place on at least two levels: the child is learning about the task, developing 'local expertise'; and he is also learning how to structure his own learning and reasoning. (Wood, 1998, p. 98)

This practice of interactive dialogue may be constrained by the cultural tradition in particular countries. In his book on how formative practices may be established in a Confucian culture, Carless (2011) points out that students in schools in China are not expected to speak up in class – the expected behaviour is to be passive and

obedient. It follows that formative assessment through oral dialogue cannot be established, although it can be developed through dialogue in writing.

It is hard to evaluate the successes of the eight projects reported here, because none of them uses direct evidence derived from observations of classroom dialogue. In the work in Sweden, all of the schools reported, in feedback given to students, an increase in pedagogical discussion and an increase in assessment for learning practices in the classroom, so that students engaged in 'dialogue around documentations of learning' (Jonsson et al., 2015). However, no detail is given about what happened in whole-class interactions, and there are no actual examples of interactive dialogue. Similar positive evidence is reported for the Canadian initiative: there, observations in the classrooms of 18 teachers were made, but direct examples of these observations are not presented and there is no discussion of the improvements achieved by the project in the quality of class discussions.

In the Singapore paper, inconsistency between teachers' conceptions of assessment and actual classroom practice is reported, but no direct evidence, of what actually happened in these classrooms, is reported (Ratnam-Lim & Tan, 2015). The USA paper goes further in listing several techniques to enhance and support pupils' contributions to classroom discussions. These were set in the more general framework proposed by Wiliam and Thompson (2007), in which 'Engineering effective classroom discussions' and 'Providing feedback that moves all learners forward' were identified as two of the key elements in the implementation of assessment for learning. The obstacles encountered in Trinidad and Tobago seem to reflect the cultural obstacles to classroom dialogue reported by Carless, where beliefs and expectations of both parents and teachers are serious obstacles: indeed the conclusion there, drawing on the analysis by Carless, is that continuous assessment practices will only work in the context of 'retroactive and pro-active forms rather than interactive versions'.

In general, the absence of direct examples, and analyses, of the actual classroom dialogue achieved in innovations to introduce assessment for learning is a common feature: a notable counterexample, a report where actual transcripts are used, is the book by Heritage (2013). I emphasise this aspect of the studies reported here because there is evidence, from surveys in the UK and in the USA, of the poor quality of classroom dialogue, characterised, in Robin Alexander's terms, as involving far more of the *rote*, *recitation* and *instruction*/*exposition* styles than the *discussion* and *dialogue* styles (Alexander, 2006; Applebee, Langer, Nystrand, & Gamoran, 2003). There is also evidence that the reports by teachers of their classroom interactions are usually more optimistic than observers' reports of the same classrooms. It is notable that in the accounts given in this issue, as in most of the abundant literature on the implementations of formative assessment practices, the exploration of classroom dialogue has not been linked either to the detailed analysis by Alexander, or to the many studies of student–teacher dialogue which predated, and were not directly linked to, formative assessment issues (see e.g. Dillon, 1988; Halliday, 1993; Mercer, 2000). It may of course be impractical to include a representative sample of transcripts of dialogue in a research paper, but there are indicators of the quality of dialogue which could be reported. Examples are: the ratio of student talk to teacher talk, whether students' contributions are in the form of single words, or short phrases, or complete sentences, or paragraphs, and whether these contributions include such 'reasoning words' as think, because, would or should (see the examples and analyses in Dillon, 1988; Mercer, Dawes, Wegerif, & Sams, 2004).

Dialogue in writing

Whilst Alexander's statements refer only to oral dialogue, Wood's is broader in scope. The purpose of the stress on comments, rather than marks, for feedback on written work is that, if the student has to respond by amending the work in response to the comments, these can be seen as developing a dialogue in writing. Such work is in the overlap area between stages 3 and 4. Furthermore, in this activity, learners may start to reflect on the weaknesses in their initial efforts. The learning aims here may be expressed as developing the learners' capacity to reflect critically on the detailed outcomes of their own work and to take initiatives to improve it.

However, in this area, the borderline area between the formative and summative functions of assessment emerges as problematic, although more seriously in some cases than in others. There is explicit reference to Sadler's (1989) analysis in three of the studies, linked to emphasis on helping pupils to understand the learning intentions so that they can audit their own progress by reference to these. The report of the replacement, in the USA study, of comment-only marking by 'mastery grading' is a novel and significant advance. However, in the light of the more fundamental development of students' capacity to take more responsibility for their own learning, there appear signs of their resistance. A necessary condition for this development is the implementation of peer assessment by pupils, which can contribute to self-assessment. Whilst this was not even attempted in some cases (e.g. Trinidad and Tobago), in others it was resisted by the pupils. It received low pupil ratings in Sweden, likewise in the USA.

One obstacle here is that students may simply be looking for confirmation that they were either right or wrong – an expectation that may be a hangover from familiarity with practices which provide marks on their work. This outlook damages their development as learners, because it usually indicates a belief in students that they have a fixed intelligence so that there is little they can do to alter the fact that one is smart, or dumb, or somewhere in-between; the alternative belief is that you can always improve by your own efforts and so you have to take the risk of attempting, rather than avoiding, dauntingly challenging tasks. The research reported by Butler (1988), and in more extensive studies by Dweck (2000), has shown both that feedback with comments and without marks encourages the latter belief, and that those who hold such belief become more effective learners, and cope better when faced with such challenges as moving from one learning environment to another.

Collaboration in group work

A different aim can be pursued by developing students' capacity to collaborate in group work. Such work has intrinsic value, in that productive collaboration with others has its own value as a skill, and because in such collaboration there is a further opportunity for students to engage in interactive dialogue. Here, as in comment-only marking, it is more feasible – than it can be in the classroom – for every student to be directly involved. There is a further advantage in that students may speak more freely to one another than they might in the presence of the teacher. However, studies in the UK and the USA (Baines, Blatchford, & Kutnick, 2009; Dawes, Mercer, & Wegerif, 2004; Johnson, Johnson, & Stanne, 2000; Mercer et al., 2004), have shown that the benefits of such discussion can only be achieved if students are carefully trained for group work. Such training, based on Mercer's work, is reported in

the USA study. Opportunities for work of this type arise in the overlap area between stages 3 and 4.

Peer- and self-assessment

Work in groups can be used for a variety of purposes. One is to discuss issues raised in a class discussion, perhaps as an intermission to explore a problem that has been raised at the whole-class level, at the end of which each group reports on its conclusion. A different purpose is for students to explore one another's written work, perhaps after the teacher has assessed it beforehand without writing on it, with the aim of comparing strengths and weaknesses between their different attempts. Such work can help students to appraise their work in the light of the learning intentions, but appraisal may be hard to perform if learners cannot relate the general formulation of learning intentions to the concrete examples of their own and their peers' work. So guidance may be needed to help learners to use the concrete examples to arrive at a more general understanding. This work clearly belongs to stage 4 in the scheme of pedagogy, in that it can be implemented both with particular pieces of written work and with learners' appraisal of their attempts at an informal test. This last point leads on to a discussion of summative assessment.

Summative assessments as metacognitive reviews

The possibility of implementing learners' self- and peer-appraisal of an informal test extends the discussion in the previous section. A test is usually set at the end of the study of a particular topic, and can serve as an overview of the learning episode. Thus, the preparation for such a test, and then the review of the responses, of oneself and of one's peers, can check on and enhance a learner's grasp of the topic as a whole. This active involvement with tests may be taken further, by following the suggestion, from work by King (1992) and by Foos, Mora, and Tkacz (1994), that students can benefit by being required to compose questions for use in a summative test. One teacher described the use of this idea:

> More significantly, pupils' setting of their own questions has proved to be a stimulating and productive means of rounding off topics and revising their work. Answering other people's questions and discussing solutions with the whole class is a very effective way of concentrating on topics that need to be revised rather than spending time on what is already known. Students 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. (Black et al., 2003, p. 54)

The learning target for such activity can be summarised as the empowerment of learners to achieve metacognition, the power to develop a grasp of their work as a whole, seeing by reflection the connections between specific elements and the overall structure. The learning aim is also to develop each learner's capacity to achieve a broad overview of their progress and to guide their development in the light of its aims.

This aspect can be taken further if the organisation of classroom work, and of the written work linked to it, presents students with more open-ended tasks where they receive less guidance. Examples would be, in science, to formulate and then test the reasons for an observed natural phenomenon, or, in English, the task of writing a newspaper account of a public event in different genres for two different types

of publication. As students engage in such an activity, the feedback should be designed to leave enough freedom so that students gain in their confidence to build their own decisions in well-informed and thoughtful ways. A significant finding reported in the USA study was that students welcomed opportunities for metacognitive reflection and valued peer collaboration – but did not welcome invitations to engage in peer assessment.

Learning to learn

The learning criteria listed in the above sections can be seen, when taken together, to contribute to helping students to become independent, responsible and effective learners. The use, in Singapore, of the term 'holistic assessment' may be a reflection of this integrated perspective. What is not clear is whether and how achievement of this overall aim is affected if some of the practices summarised above are not attempted. I have commented above on the finding that classroom dialogue may not be possible in some cultures. There may also be omission of some practices at the discretion of teachers: the USA study reports, in its tables 6 and 7, the variations in the 'fidelity of formative assessment implementation', and a simpler report of changes in teachers' choices of activities is given in table 3 of Wiliam et al. (2004).

If this overall aim is used to guide interactions with students, then the full value of assessment practices in supporting this aim of a curriculum may be realised. However, it is a challenge to many teachers, as they have to learn to guide students with less detailed instruction, but rather with a more subtle guidance which pays respect to, and helps build on, their own initiatives. This aim can be summed up in the phrase 'Learning how to learn' (Black, McCormick, James, & Pedder, 2006). This should be the main aim of schooling, because any specific skills that are learnt at school are likely to become obsolete within a short time – so that the only skill which will have permanent value will be the capacity to learn new concepts and skills (see e.g. Heritage, 2013, chapter 5). One reason why work to secure this aim may not be attempted is the temptation for teachers to do too much for their pupils, as the following quotation explains:

> Educators can take over functions that learners should be doing – learning how to learn, making up their own minds, reaching personal decisions. Such imbalance ill serves learners and can be destructive to educators. There is a fine line between empowering learners as their own people and overpowering them – making them too dependent or indebted to teacher or parent. Walking this tightrope is an aspect of the educator's spiritual discipline of a balanced life. (Groome, 2005, p. 348)

Changing teachers and their students

Teachers' change

For many teachers, adopting formative assessment practices is difficult because it involves a radical change in the way in which they relate to their students and the ways they behave in the classroom. What is called for is nothing less than a change in the ways they perceive, and strive to implement, their role as teachers.

Teachers are continually faced with making instant decisions in the highly contextualised settings of the classroom so, as Schwab (1989) pointed out, they have to draw on their personal beliefs and experiences to guide these decisions. This point may be further developed by considering the following quotation from Black and Wiliam's 1998 booklet:

> *Thus the improvement of formative assessment cannot be a simple matter. There is no 'quick fix' that can be added to existing practice with promise of rapid reward*. On the contrary, if the substantial rewards of which the evidence holds out promise are to be secured, this will only come about if each teacher finds his or her own ways of incorporating the lessons and ideas that are set out above into his or her own patterns of classroom work. This can only happen relatively slowly, and through sustained programmes of professional development and support. This does not weaken the message here – indeed, it should be a sign of its authenticity, for lasting and fundamental improvements in teaching and learning can only happen in this way. (Black & Wiliam, 1998b, p. 15, authors' italics)

Each of the eight studies set up procedures to effect teacher change. Some clearly show care in setting up regular occasions to meet in groups to exchange experiences and learn from one another. The strategy in Canada of 'Instructional Rounds' was a copy, with adaptations, of the procedure used by experienced hospital doctors in training newly qualified juniors: this is ironic, because in the UK one such senior doctor reported that he drew on the publications of the KMOFAP practice to improve his practice by encouraging interactive dialogue in his instructional rounds (Caldwell, 2011). The USA initiative also used a specific strategy, one that used the general framework proposed by Wiliam and Thompson (2007), in which 'Engineering effective classroom discussions' and 'Providing feedback that moves all learners forward' were two of the key elements in the implementation. For others, the strategy does not seem clear, apart from responding to top-down decisions about what should happen – as in the Norway and Singapore examples.

The KMOFAP strategy, subsequent to the 1998 publications, was set up as a development project to extend over two years, with the first six months spent only on each teacher trying to implement some of the ideas, by personal choice from the different practices that the research suggested. Then, in the whole-day meetings of all involved held every five weeks, teachers could report back on, and discuss with others, their experiences. Only after that initial six months did each choose a more structured commitment to trial their chosen practices with one class over a whole school year. Initially, the ideas gleaned from the research review were presented as possibilities to be explored and transformed, by such explorations, into practices which were workable for them. This approach reflected another principle, expressed in the 1998 booklet:

> Teachers will not take up attractive sounding ideas, albeit based on extensive research, if these are presented as general principles which leave entirely to them the task of translating them into everyday practice – their classroom lives are too busy and too fragile for this to be possible for all but an outstanding few. (Black & Wiliam, 1998b, pp. 15–16)

Such a slow and sustained approach has not been described in all of the studies reported in this Special Issue. It seems in some, notably for Trinidad and Tobago, that the training given was based on a relatively brief exposition of the practices to be followed, whereas in the work in Norway, Sweden and in the USA, an initial two-day course was followed by exchanges and developments of experience in teacher learning communities extending over 18 months in Norway and over two years in the other two.

The opportunities for teachers to exchange experiences, often as anecdotes, were powerful drivers of the changes they achieved. In her paper on teacher change in the KMOFAP project, Harrison stated:

> … but in sharing anecdotes with like-minded peers, our teachers developed a sense of validation and acceptance that spurs the lonesome classroom teacher to persevere with their ideas once back in their schools and just develop their sense of self-as-teacher within this community. (Harrison, 2005, p. 261)

Another feature of the approach in the KMOFAP project was the variety in the patterns of change of different teachers, which were more striking than the effects of the variability across schools, which was a problem for Norway (Hopfenbeck, Flórez Petour, & Tolo, 2015). This finding was discussed in chapter 6 of Black et al. (2003), which included accounts by three teachers, each of about two pages, of their personal experiences of change over the two years involved. The paper by Lee and Wiliam (2003) analyses the changes in more detail, with particular attention to detailed accounts of the changes over time in the work of two more teachers in the same project, whilst the brief report by Eggen of her study of teacher change (pp. 111–113 in Baird, Hopfenbeck, Newton, Stobart, & Steen-Utheim, 2014) describes how four teachers differed in that 'they construct and re-construct different identities ideologically and epistemologically'.

The fact that many of the teachers did achieve changes that they valued, and which could be attested by interviews and by observation of their classrooms by members of the project team, raised the question: Which aspects of the project's strategy supported the varied teacher changes? Lee and Wiliam's answer was to identify six aspects. First, presenting teachers with *credible evidence*. Then giving them some *practical ideas* with which to start their own explorations, accompanied by the arrangement of meetings in which they could give one another *mutual support*. Explicit report of this feature is reported in the work in Canada where, as school cultures came to embody assessment for learning, the level of conversations amongst teachers was enhanced. This same consequence was identified, in KMOFAP, as the basis for the fourth productive aspect, which was that this supportive environment did encourage, for teachers, both exchanges of details and *reflection* about their actions. A fifth aspect was the *time allowed*: at the end of the first year of the project there were only modest changes in the teachers' actual classroom practices, yet without any change in the actions of the project team, radical changes did start to appear during the second year. The final aspect was *flexibility,* in that from the outset teachers were encouraged to make their own choices, for their own explorations, from the menu of possible changes which the project presented to them.

Students' change

The above features of teachers' change cannot be understood without a parallel analysis of the changes in their students. Perrenoud emphasised the importance of this connection as follows:

> Every teacher who wants to practise formative assessment *must reconstruct the teaching contracts so as to counteract the habits acquired by his pupils.* (Perrenoud, 1991, p. 92; author's italics)

One teacher's account of his work of re-construction was quoted in Black et al. (2003):

> It became obvious that one way to make a significant sustainable change was to get the pupils to do more of the thinking. I then began to search for ways to make the learning process more transparent to the pupils. Indeed I now spend my time looking

> for ways to get pupils to take responsibility for their learning at the same time making the learning more collaborative. (pp. 94–95)

A school leader in one of the other project schools emphasised this feature with a broader view of the role of his school:

> [It's essential] that we have a greater emphasis on children's learning, that we are supporting learning far more than we are doing at the moment. I don't think that we do it particularly well. Individuals do but I don't think that we are using our assessment to progress learning. It doesn't happen overnight. So if you are saying 'what do I want in five years time?' – ideally it's that all staff are using assessment as a tool to develop children's learning. (Black et al., 2003, p. 104)

The shadow of high-stakes summative assessments operates as an obstacle through students as well as directly for teachers. For example, in Sweden, when final marking was imminent, students wanted opportunity to 'fix things' and improve their mark, and it also seemed that there was a 'fixing' mentality throughout the progress of the innovation.

Conclusions

It is not possible to summarise the diverse accounts presented in the eight papers of this Special Issue. It may be easy to say that any particular study could have been more useful if more attention had been paid to this or that feature, and I have fallen for this temptation in some of my remarks. An underlying difficulty, for any innovation, is that choices have to be made between using a selection of practices which have been replicated in earlier studies, or exploring new methods. Yet, there is no common and agreed formulation of the fundamental issues in terms of which, to guide the choices, such practices may be contextualised and judged and key features subsequently communicated.

However, it would not be reasonable to interpret such comments as criticisms of any particular paper, for it is clear is that all eight had to operate within the unique contexts of their own country's ideologies, political systems and cultural traditions. As the analysis from Chile emphasises, any educational innovation has to be seen in terms of how it enters into the complex, dynamic and contested world of complex interactions between multiple systems. It is also noteworthy that most of the studies draw attention to the extra workload that teachers have to endure if they are involved in any innovation.

Popham, in his foreword to the book by Heritage (2013, p. vii), used the optimistic title 'Dualism's dividends' to draw attention to a problem that underlies all innovations of the type that are explored here, namely that academics may increase, and report on, knowledge about how things work, but teachers are the practitioners who actually make things work. Academics in education have to find ways to build fruitful interactions between their world and the world of practising teachers if they are ambitious to explore, and to learn how to implement, the potential benefits of their work.

As an academic, I have tried, in the last three sections of this article, to report on 'how things work', by proposing a framework in which the well-known features of formative assessment may be more fundamentally understood. The sections on Assessment and Pedagogy, and on Assessment and Learning, have two purposes; the first being to show how to situate assessment within the overall framework of

pedagogy, and the other being to make clear that the value of assessment for learning lies in the ways in which it can contribute to the main aim of education, which is to develop in students the capacities of independent, effective and responsible learning. I hope that these two, taken together, may serve both as a critique of the underlying philosophy of any particular project that has aimed to secure the benefits of formative assessment, and as a guide to the design of any project that aims to do so in the future. The section on the learning of teachers and students adds a further and essential dimension to this framework, one that deserves more attention if 'dualism's dividends' are to be secured.

Note

1. For three of the eight papers, the title names the country; for the others, I refer to Wylie and Lyon as the USA, Ratnam-Lim and Tan as Singapore, Flórez Petour as Chile, DeLuca et al. as Canada and Hayward as Scotland.

Notes on contributor

Paul Black is Professor Emeritus at King's College London. His main contributions have been to curriculum development in science and in technology, and to assessment research. He chaired the government's 1998 Task Group on Assessment and Testing and also served on advisory groups of the USA National Research Council. He was a member of the Assessment Reform Group. His recent work with the King's Assessment for Learning Group has been concerned with formative and summative assessments by teachers.

References

Alexander, R. (2006). *Towards dialogic thinking: Rethinking classroom talk* (3rd ed.). York: Dialogos.

Alexander, R. (2008). *Essays in pedagogy*. Oxford: Routledge.

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal, 40*, 685–730.

Baines, E., Blatchford, P., & Kutnick, P. (2009). *Promoting effective group work in the primary classroom*. London: Routledge.

Baird, J.-A., Hopfenbeck, T. N., Newton, P., Stobart, G., & Steen-Utheim, A. T. (2014). *Assessment and learning: State of the field review*. Oxford: Oxford University Centre for Educational Assessment.

Black, P. (2013). Formative and summative aspects of assessment: Theoretical and research foundations in the context of pedagogy. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 167–178). Thousand Oaks, CA: Sage.

Black, P., Harrison, C., Hodgen, J., Marshall, M., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice, 18*, 451–469.

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

Black, P., McCormick, R., James, M., & Pedder, D. (2006). Learning how to learn and assessment for learning: A theoretical inquiry. *Research Papers in Education, 21*, 119–132.

Black, P. J., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*, 7–74.

Black, P. J., & Wiliam, D. (1998b). *Inside the black box: Raising standards through classroom assessment*. London: GL Assessment.

Black, P., & Wiliam, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal, 29*, 623–637.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology, 58*, 1–14.
Caldwell, G. (2011). Whatever happened to apprenticeship learning? *The Clinical Teacher, 8*, 272–275.
Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. Abingdon: Routledge.
Dawes, L., Mercer, N., & Wegerif, R. (2004). *Thinking together: A programme of activities for developing thinking skills at KS2*. Birmingham: Questions Publishing.
De Lisle, J. (2015). The promise and reality of formative assessment in a continuous assessment scheme: The case of Trinidad and Tobago. *Assessment in Education: Principles, Policy & Practice*, *22*, 79–103.
DeLuca, C., Klinger, D., Pyper, J., & Woods, J. (2015). Instructional rounds as a professional learning model for systemic implementation of assessment for learning. *Assessment in Education: Principles, Policy & Practice*, *22*, 122–139.
Dillon, J. T. (1988). *Questioning and discussion: A multi-disciplinary study*. New York, NY: Ablex.
Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality and development*. Philadelphia, PA: Psychology Press.
Flórez Petour, M. T. (2015). Systems, ideologies and history: A three-dimensional absence in the study of assessment reform processes. *Assessment in Education: Principles, Policy & Practice*, *22*, 3–26.
Foos, P. W., Mora, J. J., & Tkacz, S. (1994). Student study techniques and the generation effect. *Journal of Educational Psychology, 86*, 567–576.
Groome, T. H. (2005). *Educating for life*. New York, NY: Crossroad.
Hallam, S., & Ireson, J. (1999). Pedagogy in the secondary school. In P. Mortimore (Ed.), *Understanding pedagogy and its impact on learning* (pp. 68–97). London: Paul Chapman.
Hallam, S., Kirton, A., Peffers, J., Robertson, P., & Stobart, G. (2004). *Evaluation of project 1 of the assessment is for learning development program: Support for professional practice in formative assessment. Final report*. Retrieved December 13, 2014, from http://www.scotland.gov.uk/library5/education/ep1aldps-00.asp
Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and Education, 5*, 93–116.
Harrison, C. (2005). Teachers developing assessment for learning: Mapping teacher change. *Teacher Development, 9*, 255–264.
Hayward, L. (2015) Assessment is learning: The preposition vanishes. *Assessment in Education: Principles, Policy & Practice*, *22*, 27–43.
Heritage, M. (2013). *Formative assessment in practice: A process of inquiry and action*. Cambridge, MA: Harvard Education Press.
Hopfenbeck, T. N., Flórez Petour, M. T., & Tolo, A. (2015). Balancing tensions in educational policy reforms: Large-scale implementation of assessment for learning in Norway. *Assessment in Education: Principles, Policy & Practice*, *22*, 44–60.
Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). *Cooperative learning methods: A meta-analysis*. Minneapolis: University of Minnesota.
Jonsson, A., Lundahl, C., & Holmgren, A. (2015). Evaluating a large-scale implementation of assessment for learning in Sweden. *Assessment in Education: Principles, Policy & Practice*, *22*, 104–121.
King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist, 27*, 111–126.
Klenowski, V., & Wyatt-Smith, C. (2014). *Assessment for education: Standards, judgement and moderation*. London: Sage.
Lee, C., & Wiliam, D. (2003). Studying changes in the practice of two teachers developing assessment for learning. *Teacher Development, 9*, 265–283.
Mercer, N. (2000). *Words and minds: How we use language to think together*. London: Routledge.
Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: Ways of helping children to use language to learn science. *British Educational Research Journal, 30*, 359–377.

Perrenoud, P. (1991). Towards a pragmatic approach to formative evaluation. In P. Weston (Ed.), *Assessment of pupils' achievement: Motivation and school success* (pp. 79–101). Amsterdam: Swets and Zeitlinger.

Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning processes. Towards a wider conceptual field. *Assessment in Education: Principles, Policy & Practice, 5*, 85–102.

Ratnam-Lim, C. T. L., & Tan, K. H. K. (2015). Large-scale implementation of formative assessment practices in an examination-oriented culture. *Assessment in Education: Principles, Policy & Practice*, *22*, 61–78.

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

Schwab, J. J. (1989). *The reflective turn*. New York, NY: Teachers College Press.

Spenceley, P. (2009, November). 10 years of assessment for learning – A personal view. *Curriculum Management Update*, 7–10.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice, 11*, 49–65.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Lawrence Erlbaum.

Wiske, M. S. (1999). What is teaching for understanding? In J. Leach & B. Moon (Eds.), *Learners and pedagogy* (pp. 230–246). London: Chapman.

Wood, D. (1998). *How children think and learn*. Oxford: Blackwell.

Wyatt-Smith, C. M., & Bridges, S. (2008). *Meeting in the middle – Assessment, pedagogy, learning and students at educational disadvantage*. Evaluation for the Literacy and Numeracy in the Middle Years of Schooling Initiative Strand A. Retrieved from http://education.qld.gov.au/literacy/docs/deewr-myp-final-report.pdf

Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: Issues of breadth and quality. *Assessment in Education: Principles, Policy & Practice*, *22*, 140–160.