

EDUCATIONAL DESIGNER

JOURNAL OF THE INTERNATIONAL SOCIETY FOR DESIGN AND DEVELOPMENT IN EDUCATION



Assessment and the Design of Educational Materials

Paul Black & Dylan Wiliam

Abstract

To link the two entities that are identified in our title, it is first necessary to establish a set of general principles in the light of which the functions of each may be evaluated. To meet this need, we suggest two different and related approaches. The first involves consideration of the link between formative assessment and theories of learning, whilst the second is concerned the ways in which summative assessments align with the aims of learning. The account here starts with discussion of the first of these, and the implications of this discussion for the practices of assessment will then be analysed. Implications for a theory of pedagogy will be explored in the course of that discussion.

A separate discussion of the implications for the design of educational materials would have to follow a similar approach, for such materials should be designed to help teachers support the learning of their pupils as effectively as possible. So the starting point for such design must be the criteria for effective learning. Therefore, in this paper we interweave discussion of the role of assessment in supporting effective learning with the role of educational materials in supporting both. In a concluding section, we summarise the practical implications for the design of materials that follow from our analysis.

Assessment and learning

Any discussion of formative assessment cannot be conducted independently of the learning that it is meant to serve ([Black & Wiliam, 1998](#)). For teachers who, for example, believe that learning history is a matter of learning facts and dates, then formative assessment would be little more than checking that the facts and dates that were to be learned had, in fact, been learned. This could involve the teacher setting tests, students assessing each other, and perhaps even peer assessment, but the nature of the formative assessment process would be driven by the view of the nature of historical thinking. On the other hand, teachers who view learning history as requiring understanding chronology, cause and effect,

the interpretation of historical sources and so on would probably employ very different kinds of assessment. The principles of formative assessment would be the same, but the kinds of assessments used, and the way they were used, would differ ([Wiliam, 2010](#)).

The main principles of learning that ought to underlie a pedagogy that incorporates formative assessment may be set out under the following sections:

Dialogue in oral discussion

Dialogue is a key feature. As Alexander has explained:

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)

In a later and more detailed exploration of this issue, he states:

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 addition, classroom dialogue serves a second role, particularly important from the point of view of formative assessment, and that is to reveal to the teacher the learners’ developing conceptions of the matter at hand. For practical subjects, such as physical education, the difficulties that a child is having may be obvious. If a child is throwing a ball with his right hand while his right foot is in front of his left, then it just looks clumsy. However, in more academic subjects, the teacher cannot peer into a learner’s head to determine what is happening. Instead, the teacher must elicit evidence to form a model of the child’s conceptions.

One particularly useful way of eliciting evidence about the child’s conceptions that integrates assessment and instructional design is through the use of “set-piece” questions in which, at a particular point in the instructional sequence, the teacher undertakes a “check for understanding” ([Hunter, 1994](#)). Of course, the best teachers have always done this, but such checks for understanding are particularly effective if they are planned into the instructional sequence, and designed as carefully as other aspects of lesson plans.

As [William \(2014\)](#) points out, such checks for understanding are of limited value if responses are obtained only from a small number of students (especially if the responses come from the most confident students). This is why it is particularly useful to design “hinge-points” into instructional sequences—points at which the teacher wants to check for understanding, either because of the amount of time since the last such check, or because the particular material being taught is known to be “troublesome knowledge” ([Perkins, 1999](#)). While there are many design requirements for the questions (or other evidence-eliciting prompts), it seems as if the following are particularly important ([Wylie & William, 2006; 2007; William, 2011](#)), listed below in order of priority:

1. The responses chosen or given by students with appropriate conceptualizations – termed “correct cognitive rules” by [Bart, Post, Behr & Lesh, \(1994\)](#) – differ from those chosen or given by students with inappropriate conceptualizations (“incorrect cognitive rules”).
2. Different incorrect cognitive rules lead to different responses

For example, [Osborne \(2011\)](#) gives the following example of a question designed to assess a higher-order thinking skill related to observation and measurement.

Janet was asked to do an experiment to find how long it takes for some sugar to dissolve in water. What advice would you give Janet to tell her how many repeated measurements to take?

- A. Two or three measurements are always enough
- B. She should take 5 measurements
- C. If she is accurate she only needs to measure once
- D. She should go on taking measurements until she knows how much they vary
- E. She should go on taking measurements until she gets two or more the same

The important point about this item is that it is highly unlikely that students with inappropriate or incomplete conceptualizations of the relevant material are likely to choose the correct response. Moreover, those with different incorrect or incomplete conceptualizations are likely to give different responses.

Of course, the teacher can never be sure that students who provide correct responses do, indeed, have the intended understanding, nor that those who do not provide correct responses do not have the intended understanding. The teacher is always attempting to construct a model of the student’s thinking, and can never be sure that the model she has constructed is a good model of the child’s thinking. As [von Glasersfeld \(1987\)](#) notes:

Inevitably, that model will be constructed, not out of the child's conceptual elements, but out of the conceptual elements that are the interviewer's own. It is in this context that the epistemological principle of fit, rather than match is of crucial importance. Just as cognitive organisms can never compare their conceptual organisations of experience with the structure of an independent objective reality, so the interviewer, experimenter, or teacher can never compare the model he or she has constructed of a child's conceptualisations with what actually goes on in the child's head. In the one case as in the other, the best that can be achieved is a model that remains viable within the range of available experience." ([von Glasersfeld, 1987](#) p. 13)

This is why dialogue is such an important part of formative assessment. With “set piece” assessment events, the student does not have the opportunity to negotiate the meaning of the assessment task, and has to respond to the best of her or his ability, whether they understand the task or not. And the teacher, in turn, is presented with evidence that needs to be interpreted without further opportunity for clarification. With dialogue, meanings can be explored and the teacher can shift from an *evaluative* role (did the student answer correctly or not?) to an *interpretive* role (what can I learn about the student's thinking by attending carefully to what she just said?). For further details on the distinction between evaluative and interpretive ways of listening to students, see [Davis \(1997\)](#).

Dialogue can be developed in oral interactions—i.e., in *classroom discussion*—or in the interactions in writing that can develop when learners are given feedback *in written work* and are expected to respond to that feedback, although the asynchronous nature of written exchanges make such exchanges less flexible as noted above. The teacher promotes an oral dialogue by setting up and steering discussions through which students are encouraged to talk about their understanding and secondly, through that teacher's formative responses to the students ideas which helps them to re-consider and expand their understanding. The interactions evoked by formative feedback are the key to developing classroom dialogue. As [William and Thompson \(2007\)](#) stated, in their analysis of the elements of the teacher's role in formative assessment, the teacher has to be engaged in “engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding” and then in “providing feedback that moves learners forward”. Such “engineering” requires well-designed questions or tasks that evoke the interest of a class and are matched to the level of understanding of the class concerned: the term “matching” implies that the activity is challenging enough, in that students have to think about adapting their existing ideas to the task, but not too far beyond their capacity to respond.

So one way in which educational materials can help is to provide teachers with examples of such questions or tasks, with advice about the stage of development for which they have been shown to be effective.

However, to conduct such a dialogue effectively, the teacher has to steer the ensuing discussion to achieve a balance between *either* controlling too closely, which might reduce the discussion to a “guess the right answer” exercise, *or* letting it range too widely so that little progress is made towards the aims of the learning. This is a delicate and skilled task, and teachers who have been accustomed to working with a “delivery” mode of learning will find it hard to relax control. Success depends on several factors. One is care in the framing of the questions, often replacing “do you agree or disagree?” with “what do you think?”. Another is to allow time for students to think about, and to formulate, their ideas before calling for contributions. Another, and more difficult one, is to react helpfully to any response, which can be difficult when unexpected and apparently stupid responses are produced. The style that is required has been summarized by a teacher as:

Students are comfortable with giving a wrong answer. They know that these can be as useful as correct ones. They are happy for other for other students to explore their wrong answers further. (Black et al., 2003 p. 40)

The principle here is expressed by another of William and Thompson’s five aspects of the teacher’s role, namely “activating students as learning resources for one another”. A common problem is that some student answers are problematic because they seem simply bizarre rather than wrong, and it can often be difficult for the teacher to make sense of the student’s response without extended discussion (see [Black & William, 2009](#)). An alternative response is to encourage the student to lay out their thinking, for example by asking “Why do you think this?”, and then to invite others to say whether they agree or have a different answer.

In the light of this, exploring the role that educational materials can play in helping teachers develop these skills has to look for more; merely setting out a set of rules will be of limited value. What we have found helpful is to present teachers with samples of real classroom dialogue, both good, bad and indifferent, recorded as written transcripts, and using these as examples to illustrate key features. Whilst there are many scholarly books and papers on analysis of dialogue, there are few that are relevant to the teachers with whom we have worked. Examples we have used are from [Black et al. \(2003\)](#) and [Dillon \(1994\)](#). In using these in in-service sessions with teachers, we have been struck by the fact that, when comparing a poor with a good example (from [Black et al. 2003](#), pp. 36-39), participants tend to focus on the teacher’s actions and hardly ever comment on

the fact that in the poor case, the students only contribute short phrases, whereas in the good example, every student contribution is in the form of a sentence, and that only in the good example do students use such reasoning words as ‘think’, ‘because’, and disagree with one another’s ideas (see also [Mercer et al., 2004](#)). To provide such resources in sufficient variety across subjects and across school grades would be a formidable task. In addition, it might be argued that these might be more valuable as resources for in-service training providers or teacher learning circles than for teachers working on their own.

Dialogue with written work

The *feedback* that teachers give *on written work* is another form of dialogue that can promote formative interaction and self-regulation, albeit within a different mode and a longer time scale. Dialogue in writing can become particularly productive when teachers compose feedback comments individually tailored to suggest to each student how his or her work could be improved, and expect the student to then do further work in response—to correct misunderstandings and to deal with other weaknesses in the work.

In providing such feedback, the teacher has to tailor the comments to the needs of the individual, so that differentiation of the feedback is essential. However, there is more involved here. The research findings of [Butler \(1988\)](#) and of [Dweck \(2000\)](#) show that the choice between feedback given as marks, and feedback given only as comments, can make a profound difference to the way in which students view themselves as learners: confidence and independence in learning is best developed by the second choice, i.e. by feedback that gives advice for improvement, and avoids judgment. Learners must believe that success is due to internal factors that they can change, not due to factors outside their control, such as innate ability or being liked by the teacher.

This distinction is neatly illustrated by the response of a 15-year-old student named Åsa, in the Swedish town of Borås. Åsa was taught both Swedish and Philosophy by the same teacher, and the teacher, after hearing about the work of Butler and Dweck, decided to give comments, but not grades, when marking Philosophy homework. Although, because of the importance of the grades in Swedish for entry to higher education, the teacher continued to give grades for Swedish homework. Åsa, in reflecting on her experiences of getting just comments in Philosophy, and comments and grades in Swedish, wrote the following:

I have gone through the comments, but when there is a grade given, you become a little blinded by it, and focus too much on it. So personally (even though I quite possibly would complain if I did not get the grade) I would prefer you not to do it, because I have noticed that I pay more attention to the comment and learn more when the grade is not written on the paper.

Thus, by taking care with the use of their feedback, teachers are paying attention, in this written mode of interaction, to that aspect of their role - quoted above from William and Thomson, as “activating students as owners of their own learning”. Dialogue in written work gives more opportunity for learners to reflect on their expressed learning, and in this respect can help students to explore a deeper aspect of learning in developing this ownership. The central issue here is expressed by [Wood \(1998\)](#) in the following extract from his book “How children think and learn”:

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)

One function of educational material is to suggest questions that are effective in provoking learners to review and reconsider their understanding – sometimes in applying it in an unusual context. Questions for use in written work should serve these functions, being more comprehensive and searching than those that may be useful in oral dialogue. Such examples may be even more helpful to teachers if data on learners’ responses to these questions can be used to alert the teacher to some of the main features of likely responses that may require particular attention in any formative feedback.

Peer- and self-assessment and review of the learning

Many teachers, and their students, see summative assessment as a judgment providing, with its inevitable marks or grades, feedback that:

- labels students as more or less capable, and is
- terminal, designed to record rather than to improve the learning achieved.

Indeed, if it enhances a competitive ethos, it may, as pointed out above, have harmful effects on learning development.

A quite different perspective is possible. A test at the end of any learning episode, could be designed not only to summarise, but also to serve as an opportunity for a review in which test results could also be interpreted in the light of the strengths and weaknesses of the learning achieved. Indeed, there should be no sharp dividing line between a piece of written homework and a short, end-of-topic, test. One way to use the opportunity for feedback that such work provides is, as already explained above, for teachers to present each student with advice on how that student might do further work to improve the response. In general, students do review their work in preparation for a test, but it has been found that many lack an effective strategy for doing this ([Black et al. 2003](#), p.53).

An unusual approach to preparation is to ask students to work in groups to invent questions which they think might appear in the coming test: this approach has been shown, in two separate studies, to enhance performance in the subsequent test ([King, 1992](#); [Foos et al., 1994](#)).

An alternative way is to require students to work in groups to assess one another's work, i.e. providing feedback to one another: this implements peer-assessment as one main way to "activate students as resources for one another". The procedure used can take different forms: teachers may assess the work beforehand but make no indication of that assessment on the work itself: in addition, they may provide students with a mark scheme, or invite them to compose their own mark scheme. Students in a group may discuss the strengths and weaknesses of one another's work, or the work assigned to each group might be the work of another group. The teacher may simply observe the groups at work and intervene when any one group encounters particular difficulty, or seems to be missing an important point.

Yet another approach is for students to complete a test individually, but then, in groups, collaborate to generate the "best composite response" by comparing and discussing their answers.

The main purpose of such work is to encourage students, through the task of inventing test questions, or mark-schemes for test questions, and through considering the feedback generated by their peer's assessments:

- to apply the criteria for the quality of achievement,
- to help them understand how their work might be improved, and
- to help deepen their understanding of these criteria by relating them to their own specific examples, through the comparisons with, and critique by, their colleagues.

The key aim however is to help every student to check and so consolidate his or her own learning, and be helped by this process to become a more effective and responsible learner in the future. Indeed, the first role of the teachers specified in Wiliam and Thompson's analysis is "clarifying and sharing learning intentions and criteria for success", and this issue is discussed in more detail and with more practical examples in chapter 3 of [Wiliam \(2011\)](#).

In the learning of a new topic, any specification to students of these intentions and criteria will usually be of limited value: to state, for example, that the aim is "to understand the meaning of momentum, its conservation and its application in linear collisions" will convey little to a learner who is meeting the concepts involved for the first time. The task of applying such statements to the concrete examples in one another's work may help the process of developing understanding, of abstract statements, by generalisations over many particular cases. The advantages of such work were seen by a teacher in the following terms:

They feel that the pressure to succeed in tests is being replaced by the need to understand the work that has been covered and the test is just an assessment along the way of what needs more work and what seems to be fine. ([Black et al., 2003](#) p. 56)

To achieve such results, the quality of the questions, and their potential to generate responses that might generate fruitful student discussions, are essential. It is normal for educational materials to provide sets of questions, but such resources might be more useful if they could be accompanied by a commentary that both reported some evidence on how they might promote reflective consideration by students, and justified their validity, i.e., identified the broader aims in the teaching of the subject that they were designed to assess. Another approach that might help teachers is to provide, for any one topic, a pair of tests that are equivalent in reflecting the aims of the learning. Pupils might then discuss one of the pair, as it stands, analyzing how it reflects the main aims of their learning, whether it is a valid test of their learning, and considering what examiners might look for in their answers—work which helps prepare them for the formal test for which the second of the pair would be used.

Group work

The several activities proposed in this section all require that students work in groups. For such work to be beneficial, research has shown both that participants have to interact in co-operation rather than in competition ([Johnson et al., 2000](#)), and that groups often fail to achieve this ([Blatchford et al., 2006](#); [Baines et al., 2009](#); [Mercer et al., 2004](#)). However, in these last three studies, it has been shown that primary-age pupils can be trained to work co-operatively and that such training can help them to make better progress in achieving the aims of the

learning. Thus, one need that educational materials can help to meet is to provide materials that teachers may use to train their students in conducting group work. An excellent example is provided in the publication entitled “Thinking Together” (Dawes et al. 2004), which is addressed to teachers of pupils aged 8 to 11. This publication sets out detailed plans for a sequence of 16 lessons: the first five of these help set up exercises in, and ground rules for, group work, and these are followed by outlines of activities in which group work is used to develop the skills listening and thinking together. The prompts and texts for the pupils to use are provided, in photocopiable form

A similar approach, albeit more comprehensive but with less detailed guidelines, is taken in a book about promoting effective groups, with sub-title “A hand book for teachers and practitioners” (Baines et al., 2009). This also sets out a schedule of training activities in group work: the 13 units dealing with such issues as ‘sensitivity, respect and sharing views’, ‘giving reasons and weighing up ideas’ and ‘decision making: consensus and compromise’: each of these has been formulated in the light of findings in the authors’ research surveys which have revealed the weaknesses in group work which they are designed to correct.

Summative assessments and learning aims

The sequence of decisions and action involved in the design and implementation of any learning programme may be set out in a simple way in the following model of pedagogy (Black, 2013, p.210):

- A. **Formulating Aims.** This is the stage of strategic decision. All that follows should relate to a clear formulation of the learning aims.
- B. **Planning Activities.** The aims are to be achieved by choosing, adapting, or inventing activities that will engage pupils, and thereby elicit responses from them which help to clarify and then extend their understanding.
- C. **Implementation.** The way in which a plan is implemented in the classroom is crucial. What is needed is formative interaction that stimulates and builds on the pupils’ contributions. This is the core activity of assessment for learning.
- D. **Review.** At the end of any learning episode, there should be a review, to check before moving on. The assessment used at this stage may be designed to be summative, but its results can also to support learning, e.g. to help all pupils, through peer marking, to develop understanding of the criteria of quality in meeting the aims. It may also help the teacher to identify a need to revisit some issues with the class as a whole.
- E. **Summing Up.** This is a more formal version of the *Review* stage: here the results may be used to make decisions about each pupil’s future work.

Whilst the picture presented in the previous section is mainly about the Review in Section D, where formative and summative functions intertwine, Section E is about a formal and terminal process. However, it should be inter-related with, consistent with and supportive of the pedagogic process as a whole. The questions, or other forms of assessment, that are central to this stage should reflect the overall aims of the learning programme. In so doing, they have an important function, one that is illustrated by the following statement:

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. (Wyatt-Smith & Bridges, 2007 p. 48)

It is attractively easy to state broad and admirable aims in setting out any presentation of a plan for learning, but this may also be indulgent in that the precise meaning of these aims may not be clear. In setting final summative assessments, these aims are made clear in the achievements that such assessments reflect and reward. The danger is that these assessments may not actually reflect, and generate evidence relevant to, the initial aims, or that the meaning of those initial aims seemed acceptable at a stage when they had not been made clear by the discipline of composing assessments that would achieve such clarification.

This is an important point because adopting particular assessments for use in determining what students have learned forces the teacher to “get off the fence”. Teachers may appear to have a reasonable consensus about what is meant by understanding the concept of conservation of energy, but when a particular assessment is adopted as a way to judge whether students have, or have not understood the concept, then any disagreements are immediately highlighted. In other words, “assessments operationalize constructs” (William, 2010 p. 259). This is why any statement of aims should be clarified and expressed at the outset by explicit specification of their meaning in the final assessments. The test has to be aligned with the aims and with the teaching – ‘teaching to the test’ is not a problem if the test embodies the important aspects of the work at hand.

Carefully-framed and tested educational materials can support all stages of this process, and the provision of ready-made summative test items is the most obvious and familiar example of such support. Such support can nevertheless be problematic. It is widely reported that teachers in several countries lack understanding and skill in composing their own tests or in evaluation of tests provided by other agencies (see, for example, Harlen, 2005, Webb, 2009). This is

due in part to the neglect of this dimension of the teacher's role in teacher training, and in part to the pressures of high-stakes accountability tests over which teachers have no control and to the demands of which they are forced to conform. Nevertheless, teachers and schools do have responsibility for the summative assessments of their students, at least in those school years when accountability tests are not imposed, and since the results in those years are used to inform decisions about students' progress and the choices to be made about each student's next steps, it is important to secure a high level of validity for these assessments. Yet a small scale study by [Black et al. \(2011\)](#) showed that some teachers use tests provided by external agencies without considering the validity of such tests and the alignment of them with their own teaching aims.

Whilst educational materials cannot on their own overcome any shortcomings in teachers' assessment skills, they might be helpful in providing information about the properties of any test materials. Providing information about reliability would be fairly straightforward. Validity is a more difficult issue ([William, 2011](#); [Newton & Shaw, 2014](#)), not least because of the different aspects that might be emphasized (e.g. content, predictive, concurrent, etc.). The most helpful would be to stress that validity is concerned with the inferences that users of the results might draw from them. Thus, if students were to use their results in a particular study to decide that they could not do well in any further study of that subject, or if those teachers who would have to teach a class of students in the next school year were to decide that only one of the topics listed in the curriculum of the year just completed would need any further attention from them, would these decisions be justified by the brief summary evidence that assessment data usually present? The value of any test materials offered to schools might be improved by attention, in their initial development and evaluation, to the uses to which the information generated by those test materials is likely to be put. This might mean that their designers would provide a commentary to make clear the aims, in terms of topics and levels of understanding which they addressed, and the different implications that they would advise teachers to consider, between students who might achieve (say) the highest level and those who might be at (say) the 50% level.

How can educational materials support teachers' assessments?

A framework of aims

Answers to this question can be considered in the light of three perspectives. The first of these is the model of pedagogy presented above. The issue here is to consider for which of the stages A to E in that model can educational materials provide support. It seems clear that for stage B, *Planning activities*, and stage E, *Summative assessments*, exemplary materials can make a direct contribution, although even in these cases attention has to be given to the contexts for which such materials were designed or evaluated. For Stage A, the *Learning aims*, the

issue may be that the aims that materials are designed to serve or that they help to clarify (as in the case of summative assessments at E clarifying the meaning of the aims) should be made clear. After all, the task of educational materials is not to determine, or even modify, the aims that teachers should adopt for their teaching. The least clear case is that of stage C, *Implementation*, because in classroom implementation the problems that might arise are unpredictable and flexible adaptation by the teacher is essential. However, exemplary materials designed to support the pre- or in-service training of teachers may have an important role here. Here the discussion has to focus more directly on the teacher's role in helping students to develop as confident and competent learners, as outlined in the 2007 analysis of [William and Thompson](#).

Thus, a second perspective, with focus on the teacher's role, emerges. This is most clearly presented by Table 1 from William and Thompson's paper.

Table 1: Aspects of formative assessment

	Where the learner is going	Where the learner is right now	How to get there
Teacher	1 Clarifying learning intentions and criteria for success	2 Engineering effective class-room discussions and other learning tasks that elicit evidence of student understanding	3 Providing feedback that moves learners forward
Peer	Understanding and sharing learning intentions and criteria for success	4 Activating students as instructional resources for one another	
Learner	Understanding learning intentions and criteria for success	5 Activating students as the owners of their own learning	

Each of the five numbered types of action has been introduced at the relevant stages in our discussions above, i.e. it has already served as a guide to the development of our argument. What should be made clear in this summary is the link of these five to the aspects of theories of learning to which they were linked. The following listing of the elements of this third perspective links each of their aspects to the components of Table 1.

Dialogue in oral mode:

this relates directly to components 1 and 2

Dialogue with written work:

this also relates to 1 and 2, but can have more emphasis on the ways in which such work encourages students' reflection on their work and thereby serves the overall aim for peers of "Understanding and sharing learning intentions and criteria for success" and so can address component 4, and insofar as feedback is given to each individual student, also help the teachers aims in component 5.

Peer- and self-assessment and review of the learning:

here component 4 is addressed more directly, whilst also contributing to number 5.

Summative assessments:

the role of these in helping all concerned to give explicit meaning to their learning intentions has already been emphasized here. For learners, their own work on summative tests, in assessing their own responses and those of peers, may contribute to developing their understanding of the meaning of the criteria. Other activities here, both in reviewing work in preparation for a final assessment and in an exercise of trying to compose a written summative test, may also contribute in this dimension.

Implications for design

So far in this paper, several practical suggestions have been made. These may be summarized as follows:

- I. Provide examples of questions or tasks that can engage students in expressing and exchanging their ideas about a phenomenon or topic.
- II. Provide samples of classroom dialogue that teachers might analyse to develop deeper understanding of their own classroom style.
- III. Give examples of various types of question that could be effective in stimulating students to review and re-consider their own understanding.
- IV. Provide summative tests, explaining the purposes for which the results of these would be valid evidence, perhaps with tests in equivalent pairs to promote predictions and/or analysis by students.
- V. Specify outlines designed to develop the skills of collaborative group work amongst students.

These are only examples, and it is clear that for most of them the materials would have to be different for each curriculum subject. A more general difference is that

one class of such materials could be designed to provide ‘off-the-shelf’ material for teachers to use within existing lesson plans – examples I and III and IV above could be in this category: in terms of the five-stage model of pedagogy proposed above, such materials would be of direct use in stage B (planning activities), in stage E (summing up) and, less directly, in helping to align the plans in stage E to the aims in stage A.

Other materials would be different in being aimed to enhance the role of the teachers in assessment, and here the guide would be the analysis of Wiliam and Thompson as set out in Figure 1. Examples II and V are in this category: material for these purposes would be best used in enhancing professional development whether through initial or in-service training, or in collaborative collaboration between teachers themselves.

All materials, whether within or cutting across these categories, must be designed, as we have emphasised in the first part of this paper, in the light of the learning aims that they are meant to serve. Moreover, even tests which could be used for ‘off-the-shelf’ purposes, might be more useful in the long-term if they contributed to, rather than replaced, each teacher’s responsibility for selecting and adapting any questions to both clarify and reflect their own aims and to serve the needs of their students. In this way, tests of type IV above could both meet a teacher’s immediate needs and contribute to the development of that teacher’s assessment skills. Such positive effects were reported in Black et al. ([2011](#), [2013](#)).

The importance of such work has been emphasized by many authors, For example, Stanley, drawing on experience in both Australia and the U.K. stated that:

Evidence from education systems where teacher assessment has been implemented with major professional support, is that everyone benefits. Teachers become more confident, students obtain more focused and immediate feedback, and learning gains can be measured. An important aspect of teacher assessment is that it allows for the better integration of professional judgment into the design of more authentic and substantial learning contexts. ([Stanley et al., 2009](#) p. 82)

Whilst that statement could be read to apply mainly to summative testing, the broader outlook that the authors had in mind is clearly stated in an earlier extract from the same article:

... the teacher is increasingly being seen as the primary assessor in the most important aspects of assessment. The broadening of assessment is based on a view that there are aspects of learning that are important but cannot be adequately assessed by formal external tests. These aspects require human judgment to integrate the many elements of performance behaviours that are required in dealing with authentic assessment tasks. (Stanley et al., 2009 p.31)

This last extract serves our purpose in presenting this article. The point is that attention to assessment purposes should not be seen as a refinement in the design of educational materials, but rather as fundamental if they are to serve the most important aims of education.

References

- Alexander, R. (2006). *Towards dialogic thinking: Rethinking classroom talk*. York, UK: Dialogos
- Alexander, R. (2008). *Essays in pedagogy*. Abingdon, UK: Routledge.
- Baines, E., Blatchford, P. & Kutnick, P. (2009) *Promoting effective group work in the primary classroom* London, UK: Routledge
- Bart, W. M., Post, T., Behr, M. J., & Lesh, R. (1994). A diagnostic analysis of a proportional reasoning test item: An introduction to the properties of a semi-dense item. *Focus on Learning Problems in Mathematics*, 16(3), 1-11.
- Black, P. (2013). Pedagogy in theory and in practice: Formative and summative assessments in classrooms and in systems. In D. Corrigan, R. Gunstone & A. Jones (Eds.), *Valuing assessment in science education: Pedagogy, curriculum, policy* (pp. 207-229). Dordrecht, Netherlands: Springer.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice*, 5(1), 7-73.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31.
- Black, P., Harrison, C., Hodgen, J., Marshall, B. & Serret, N. (2013). *Inside the black box of assessment: Assessment of learning by teachers and schools* London, UK: GL Assessment.

- 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*, 18(4), 451-469.
- Black, P., Harrison, C., Lee, C., Marshall, B. & William, D. (2003). *Assessment for Learning – putting it into practice*. Buckingham, UK: Open University Press.
- Blatchford, P., Baines, E., Rubie-Davies, C., Bassett, P., & Chowne, A. (2006). The effect of a new approach to group-work on pupil-pupil and teacher-pupil interactions. *Journal of Educational Psychology*, 98(4), 750-765.
- 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.
- Davis, B. (1997). Listening for differences: an evolving conception of mathematics teaching. *Journal for Research in Mathematics Education*, 28(3), 355-376.
- Dawes, L., Mercer, N., & Wegerif, R. (2004). *Thinking Together: A programme of activities for developing thinking skills at KS2*. Birmingham, UK: Imaginative Minds.
- 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.
- Foos, P.W., Mora, J.J. & Tkacz, S. (1994) Student study techniques and the generation effect. *Journal of Educational Psychology*, 86(4): 567-76.
- von Glasersfeld, E. (1987). Learning as a constructive activity. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Harlen, W. (2005). Trusting teachers' judgement: research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education*, 20(3), 245-270.
- Hunter, M. (1994). *Enhancing teaching*. New York, NY: Macmillan.
- Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). *Cooperative learning methods: A meta-analysis*. Minneapolis, MN: University of Minnesota.

- King, A. (1992) Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist*, 27(1), 111-26.
- 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(3), 359-377.
- Newton, P. E. & Shaw, S.D. (2014). *Validity in educational and psychological assessment*. London UK : Sage
- Osborne, J. (2011). *Why assessment matters*. Paper presented at the Annual conference of the Science Community Representing Education, London, UK. Retrieved on June 30, 2014 from <http://www.score-education.org/media/6606/purposejo.pdf>.
- Perkins, D. (1999). The many faces of constructivism. *Educational Leadership*, 57(3), 6-11.
- Stanley, G., MacCann, R., Gardner, J., Reynolds, L. & Wild, I. (2009). *Review of teacher assessment: What works best and issues for development*. Oxford, UK: Oxford University Centre for Educational Development.
- Webb, D. C. (2009). Designing professional development for assessment. *Educational Designer*. 1(2).
- William, D. (2010). What counts as evidence of educational achievement? The role of constructs in the pursuit of equity in assessment. In A. Luke, J. Green & G. Kelly (Eds.), *What counts as evidence in educational settings? Rethinking equity, diversity and reform in the 21st century* (pp. 254-284). Washington, DC: American Educational Research Association.
- William, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree.
- William, D. (2014). The right questions, the right way: What do the questions teachers ask in class really reveal about student learning? *Educational Leadership*, 71(6), 16-19.
- William, 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 Associates.
- Wood, D. (1998). *How children think and learn*. Oxford, UK: Blackwell

- Wyatt-Smith, C., & Bridges, S. (2007). *Meeting in the middle: Assessment, pedagogy, learning and educational disadvantage (Final evaluation report for the Department of Education, Science and Training on Literacy and numeracy in the middle years of schooling)*. Brisbane, Australia: Queensland Government Department of Education, Training, and the Arts.
- Wylie, E. C., & Wiliam, D. (2006). *Diagnostic questions: is there value in just one?* Paper presented at the Annual meeting of the National Council on Measurement in Education, San Francisco, CA.
- Wylie, E. C., & Wiliam, D. (2007). *Analyzing diagnostic questions: what makes a student response interpretable?* Paper presented at the Annual meeting of the National Council on Measurement in Education, Chicago, IL.

About the Authors

Paul Black worked as a physicist for twenty years before moving to a chair in science education. He was Chair of the Task Group of Assessment and Testing, which advised the UK Government on the design of the National Curriculum assessment system. He has served on three assessment advisory groups of the USA National Research Council, as Visiting Professor at Stanford University, and as a member of the Assessment Reform Group. He was Chief Examiner for A-Level Physics for the largest UK examining board, and led the design of Nuffield A-Level Physics. With Dylan Wiliam, he produced the review of research on formative assessment that sparked the current realization of its potential for promoting student learning.

Dylan Wiliam is Emeritus Professor of Educational Assessment at the Institute of Education, University of London where, from 2006 to 2010 was its Deputy Director. In a varied career, he has taught in urban public schools, directed a large-scale testing program, served a number of roles in university administration, including Dean of a School of Education, and pursued a research programme focused on supporting teachers to develop their use of assessment in support of learning.

Black, P., Wiliam, D. (2014) Assessment and the Design of Educational Materials.
Educational Designer, 2(7).
Retrieved from: <http://www.educationaldesigner.org/ed/volume2/issue7/article23/>

© ISDDE 2014 - all rights reserved