PAUL BLACK

22. ASSESSMENT IN SCIENCE EDUCATION

INTRODUCTION

Two functions of assessment, the formative and the summative, are often seen as independent of one another, and even in conflict. This chapter will first present an overview of the roles on assessment in supporting learning and in the broader context of a model of pedagogy. The account of formative assessment will argue that it is a central feature of effective teaching, and will stress the importance of feedback in guiding teachers in the effective implementation of plans to engage pupils in their learning of science. It will also be argued that science education can contribute to the broader development of pupils as effective learners. The links between, and distinct purposes of, informal and formal summative assessment will be explored, stressing the specific role of formal summative assessment in guiding learners in making decisions about future stages in their study of science. Overall, the analysis presents opportunities and challenges to both teachers and their pupils.

In 1998 my colleague Dylan Wiliam and I published an article entitled *Assessment and classroom learning* which reviewed about 250 research papers about the topic and drew some general conclusions from them (Black & Wiliam, 1998a). One overall conclusion, based on results of a diverse set of research studies which focussed on the feedback that teachers gave to their pupils, showed that attention to this aspect of teaching produced a significant increase in their pupils' subsequent test achievements. Over the next sixteen years, this article, and a short booklet for teachers summarising its findings (Black & Wiliam, 1998b), have been cited by many authors, and has been a starting point of further work by Dylan Wiliam and myself, with other colleagues at King's College, and by research workers in many countries. In consequence, our understanding of this aspect of teaching had evolved, both with respect to its theoretical implications and to its practical applications. The latter will serve as the starting points for the presentation in this chapter, but as each practical activity is explained, the ways in which it relates to fundamental implications for theories of learning and of pedagogy will be explained.

The key concept at the heart of this work has been formative assessment. This may be defined in the following way:

> An assessment activity can help learning if it provides information to be used as feedback, by teachers, and by their students, in assessing themselves and each other, to modify the teaching and learning activities in which they are engaged.

K. S. Taber & B. Akpan (Eds.), Science Education, 295–309.
© 2017 Sense Publishers. All rights reserved.

Three features of this definition require emphasis. One is the term ‘feedback’, which implies some interaction between teacher and pupils.[1] The second is the inclusion of both teachers and pupils in the use of feedback, implying that both teachers and pupils learn from one another. The third is the term ‘modify’, which implies that information derived in interactive feedback can lead the teacher to alter, even abandon, the original design of the teaching work because the feedback has shown that it assumes an existing understanding by pupils which they have not achieved: the approach taken has to be changed.

A direct exploration of these features is presented in the next section on the ways in which assessment can support learning. The subsequent section explores the question of how assessment fits in to the broader context of an overall model of pedagogy. Two further sections examine implications, in turn for teachers as assessors, and for pupils as learners. A closing section summarises the principles involved and the personal challenges for change that they present.

ASSESSMENT SUPPORTING LEARNING

Feedback in Oral Dialogue

A teacher ought to start any lesson on a new topic with a question designed to explore what the pupils already know and understand about the topic: teachers sometimes find that their pupils know far less than they had anticipated, but at other times or with a different class, that pupils have everyday experiences and ideas about the topic on the basis of which they have built ideas or practices which may appear to contradict the more powerful concepts about which they ought to learn.

Here, for example, is a quotation from a teacher who had taken seriously the idea that she must start by helping pupils to explain the ideas that they already had about a new topic:

- *Questioning*
  My whole teaching style has become more interactive. Instead of showing how to find solutions, a question is asked and pupils given time to explore answers together. My Year 8 [i.e. 12–13 year olds] target class is now well-used to this way of working. I find myself using this method more and more with other groups
- *No hands*
  Unless specifically asked pupils know not to put their hands up if they know the answer to a question. All pupils are expected to be able to answer at any time even if it is an ‘I don’t know’.
- *Supportive climate*
  Pupils are comfortable with giving a wrong answer. They know that these can be as useful as correct ones. They are happy for other pupils to help explore their wrong answers further. (Black et al., 2003, p. 40)

This teacher found that giving pupils time to explore answers together was essential – research has shown that many teachers expect pupils to answer within less than one second and that if no-one does so, they answer their own questions or ask another one. But if a thoughtful answer is expected, pupils needs time to think and to compose ways of expressing their thoughts. The teacher had also found that where pupils were expected to volunteer an answer, the same pupils would put up their hands each time, whilst many others would not do so through lack of time to think, or through fear of producing a 'wrong answer'. To deal with such problems, she made her class realise that she wanted to know what they thought. So, for example, if a teacher were starting lessons in science about light, instead of telling the class the laws of reflection and refraction, he or she might ask:

> *Which is the odd one out – piece of white paper, mirror, picture, television? Why?*

There is no 'right answer' to this question, but in arguing about alternative answers the pupils will reveal their existing ideas about the nature of light. In this example, it is clear that if pupils are to learn effectively, they have to express and share any naïve conceptions or misconceptions which they already have and to understand how the scientific principles form a more coherent and effective way of dealing with problems. Other examples of such 'open' questions are:

> *If you keep a drink with ice cubes in a thermos flask, do you need to leave room for the ice cubes to melt?*

or

> *What are the similarities, and what are the differences, between combustion and respiration?*

The aim of such questions is to get pupils talking about the subject of the lesson. If such talking is to be encouraged, the teacher has to encourage it: so to ignore a strange response, or to state that it is wrong, is not helpful – a far better response might be 'Why do you think that?', and then to accept any explanation and ask the class 'Does anyone have a different idea?' The teacher's task here is a delicate one, for on the one hand the discussion must not be allowed to wander too far away from the main aim of the lesson, whilst on the other hand pupils must be helped to understand the new ideas with which the lesson is challenging them. A more detailed exploration of classroom dialogue in relation to scientific inquiry has been reported by Ruiz-Primo and Furtak (2006).

Some might object that there are better ways of using the time that such discussions will require. The issue here is about the value of talk in developing learning. Alexander (2008) expresses a clear view on this issue

> 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. (p. 9)

Most teachers will realise, on reflection, that they learn a great deal from talking with colleagues about problems that arise in their own work.

In summary, by encouraging pupils' involvement in classroom dialogue, teachers pursue two aims. One is to match the work to the capacity of the pupils to develop their understanding of the concepts and methods involved. The second, and more general, aim is to develop their pupils' skills in learning, through discussion, to engage in reasoning about their own ideas and about those of fellow-pupils.

Feedback in Written Dialogue

The above account has focussed on oral dialogue. There is also the possibility of dialogue in writing, i.e. of interaction by the exchange of written work. When teachers set pupils a task to write on their own, either as homework or as work in class time, the teacher collects and studies that work and returns it to each pupil with feedback. Such feedback can take the form of a mark (or grade), or of some comments about the work, or of both marks and comments. Research studies have shown that the provision of comments can improve pupils learning, whereas the provision of marks does not do so. The same work has also shown that if both marks and comments are provided, the positive effects of the learning are not produced (Butler, 1988). Pupils regard the marks as a judgment of them and will not think of the feedback as a help to their learning unless they are only given helpful comments. A teacher who took this evidence to heart reported as follows:

> Students do work on targets and corrections more productively if no grades are given. (The researcher) observed on several occasions how little time students spend reading my comments if there were grades given as well. (Black et al., 2003, p. 43)

This teacher was using each pupil's work, with her comments about how to improve it and the subsequent efforts of the pupil to improve the work, as a learning dialogue: this made full use, of the effort invested in setting the work, in the pupil's time spent in doing it, and in the teacher's time in reading it to give feedback. By contrast, simply putting down and recording a mark usually makes little contribution to pupils' learning. Some teachers have kept a record on the comments which they made and of each pupil's responses: they found that such a record gives a far better guide for reporting on each pupil's progress than a set of marks, both for reporting to that pupil's parents and for reporting to the school management.

In summary. feedback in writing adds to the advantages for oral feedback. It is more feasible to ensure that each individual pupil receives guidance about individual

problems, and by being at a slower pace it helps pupils to reflect on their own work and to think critically about the quality of their own arguments. However, there is a more fundamental issue involved here.

Feedback can Promote Confidence in Learning

If pupils are given marks repeatedly, whether on written work, or on class tests, or on both, they come to regard these as judgments of themselves. The research by Butler et al. (1988) and more extensive work by Dweck (2000) shows that feedback can develop one or other of two opinions that pupils form about themselves as learners. Some see the feedback as a way to compare themselves with others and for a view that they are either inherently of high intelligence or inherently of poor intelligence, and that there is little that they can do about this. Others see any feedback as a stimulus to improve and believe that they can improve by their own efforts. If they form the first of these two 'mind-sets' they become reluctant to take risks, failures simply damage self-esteem, and they react badly to new challenges, such as those which arise when they change schools or go on to higher education. If they form the second, they become willing to take on new challenges and to learn from failure, and they cope more positively with the challenges presented by new environments or new learning requirements. Such work has also shown that feedback given as marks can make pupils develop the first mind-set, whilst feedback given only as comments to guide improvement develop the second. It is because of such findings that Dweck advises that parents and teachers should never praise a child, but rather that they should praise what the child has achieved, i.e. to say "you are a clever child" encourages the inherent fixed ability view, whereas to say "your answer to this question was very good" encourages the confidence that one can, with effort, overcome difficulties. Teachers, parents, and the pupils find it difficult to accept a change from marks with or without comments, to comments only – but it has been achieved. To report one teacher's experience:

> Students are not good at knowing how much they are learning, often because we as teachers do not tell them in an appropriate way
>
> When asked by a visitor how well she was doing in science, the student clearly stated that the comments in her exercise book and those given verbally provide her with the information she needs. (Black et al., 2003, p. 46)

In summary, to replace the frequent provision of marks, grades and rank-order lists by comments aligned to the need of each individual pupil is important for two reasons. At one level, they can reflect upon their own first attempts and learn how to improve on them. At a deeper level pupils can be moved away from seeing feedback as a judgment on their 'innate ability' to seeing it as guide which encourages the belief they can improve performance by their own thoughtful efforts.

Peer-assessment by Pupils

If it is assumed that the teacher is the only person in the classroom who can help pupils with their learning, then the time that can be given in interactive dialogue with any one pupil is obviously limited. However, pupils can help one another, and often do so. Teachers can encourage such peer interaction. One example was that the teacher looked, after a lesson, at the written work pupils had handed in, but gave their work back to the pupils at the start of the next lesson without writing anything in their books. Pupils were then asked to work in groups of about five, asked to circulate one another's work around the group, and then discuss the strengths and weaknesses of each example. The main aim was that this would involve every pupil in discussing the differences in quality between their own work and that of fellow pupils. One teacher reflected on the value of such work as follows:

> We regularly do peer marking—I find this very helpful indeed. A lot of misconceptions come to the fore and we then discuss these as we are going over the homework. I then go over the peer marking and talk to pupils individually as I go round the room. (Black et al., 2003, p. 50)

This practice helped that teacher to decide where she might best spend her time, whilst ensuring that all pupils would be involved. The involvement of pupils in comparing one another's work has two advantages. One is that any opinion will have to be related to the criteria by which such work should be assessed: as pupils come to think about the meaning of such criteria in relation to the concrete examples from their own and one another's work, their understanding of the aims of their learning work will be improved: this matters because pupils can only enhance the quality of their own work if their efforts are guided by their understanding of the aims of the work and of the criteria by which achievement of such aims may be judged.

A related advantage is that, by seeing how their work compares with that of peers and is evaluated by them, pupils may become better at reflecting on their own work and at estimating its strengths and weaknesses. The importance of this feature was expressed by a psychologist as follows:

> 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)

In summary, peer-assessment can contribute to the development of self-assessment, which is an important ability in the development of pupils as mature and independent learners, able to structure their own learning and reasoning.

Effective Group Work

However, peer-assessment can only be effective if pupils are able to work co-operatively in groups. Studies of group work in schools show that this is not always achieved, and that if the pupils see the group as an arena for competition then the work will not improve learning. Other research has confirmed this finding and has been linked to studies of ways to train pupils to co-operate more effectively (Blatchford et al., 2006; Mercer et al., 2004). Mercer's training was based on requiring pupils to respect four rules: these were that no pupil should dominate the conversation and that no pupils should remain quiet for too long, that groups must work under pressure to achieve consensus and to report that consensus to the whole class, that all contributions must be treated with respect, and that any pupil who states an assertion or a contradiction must be required to give a reason for their statement. This training produce two important effects: one was that after the training such words as 'think', 'should' and 'because' occurred three times more frequently in the group discussion than they had occurred in the same groups before the training. Another was that groups so trained gained higher scores in subsequent test of the topics discussed than comparable groups who had not been trained.

These studies have also shown that the number in a group should be more than two, because when most pupils talk in pairs they tend to agree with one another too readily, but limited to about five to ensure that all members have opportunities to participate. It also helps if there is a range of achievement history within a group so that a range of different levels and types of response will arise, but that this range must not be so large that those at one extreme cannot communicate with those at the other extreme. It has also been found helpful, in mixed gender classes, to have a mix of genders in each group. Finally, pupils should not be left to choose existing friends to form a group: as one pupil put it "I don't argue with my best friend".

In summary, effective group discussions can develop further the benefits of feedback on regular written home-work or class-work. More fundamentally, it can help pupils develop their own 'mental self-regulation' whilst also making them more capable of contributing to, and learning from, group work with their peers.

Formative use of Summative Assessments

In addition to their use in enhancing pupils' learning from their regular written work, group discussions can also be used to help pupils reflect on their performance in informal testing. Any assessment may be used either to serve as a guide to how learning may be improved, or as a summary of the quality of the learning achieved at a given stage. Thus a question asked informally in a classroom may serve as a confirmation that there is no need to repeat an explanation, whereas a formal test set at the end of a year may serve to guide teachers, or parents, or school principals, about the best level of study for a given pupil in the next year. In the first case the

purpose is formative, in the second case it is summative. However, it is possible for a test to serve both purposes.

A short test at the end of the teaching of a topic may help teachers to decide that they can move on to the next topic, or it may indicate that there is a need to attend to one gap in understanding because this gap will be an obstacle to the next stage in learning. In fact, an informal end-of-topic test can be treated in the same way as a written homework task – each pupil can be given feedback about how to improve some answers, or the pupils' answers can be explored by peer-assessments in groups. At the same time, each teacher can note the overall mark of each pupil for future summaries of that pupil's work. As for any written task, the work invested in the setting and marking a test might be used more productively if the test is not set when the topic work is due to end, but a short period time before that so that there is time to use the test results to repair faults that the test has revealed. A further shift in perspective is to regard such a test as a guide to learning rather than as a final judgment. One teacher reflected on the effects of using this approach in the following way:

> After each end of term test, the class is grouped now to learn from each other. [The researcher] has interviewed them on this experience and they are very positive about the effects. Some of their comments show that they are starting to value the learning process more highly and they appreciate the fact that misunderstandings are given time to be resolved, either in groups or by me. 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)

That teacher's pupils had come to see that one valuable function of a test is to serve as a review for each of them of achievements at the end of a phase of learning. This change in the attitudes towards, and best use of, summative testing does not of course mean that all tests must be solely formative. It does imply that a test can be helpful to learning in more than one way but it does not imply that the high-stakes summative test have no other valuable function.

In summary, an overall review at the end of work on a topic can be particularly useful to pupils if they can be involved in using it to check for any faults in their understanding of that work and to correct them.

ASSESSMENT IN A MODEL OF PEDAGOGY

To implement each of the learning activities discussed above, teachers will have to insert them into their teaching plans and thereby enrich these plans. So the question to be considered in this section is how such activities should be located within a teaching plan.

It is obvious that dialogue at classroom level is located in the everyday work of classroom teaching. However, interactive dialogue can only work well if it arises out of an activity, often a problematic question, which can both engage the attention of the pupils and present them with a challenge about which they can think and talk. Thus prior to the classroom dialogue, there is the teacher's work of preparing each lesson.

The lesson-planning stage involves thought about both the short term-purpose, of evoking an interactive dialogue, and the longer-term purpose of developing particular aspects of the development of the pupils as effective learners of science. Such development will be a marriage of the understanding of key concepts of a science topic with the aim of improving pupils' capacity to reflect on and be critical of their thinking, and to engage in discussion with others. This implies that the planning stage will interact with a prior stage of formulation of the general aims, for these aims should guide the teaching of science.

The three steps outlined above, of classroom interaction, lesson planning, and formulation of aims, form, in reverse, a natural time sequence, i.e. aims *first*, leading to lessons plans *second*, leading to class room activity *third*. However, the activities of summative testing come after these: a teacher may set an informal summative test as a review of the work in a set of lessons, which will help to consolidate the classroom work and may lead back to more classroom work if the test exposes that some important aim has not been achieved. This testing activity is closely linked to the classroom aims which it san serve to strengthen and complete. It can also be seen as a *fourth* step in the model of pedagogy.

This model cannot be complete without the addition of a *fifth* stage, which is the formal summative test. What distinguishes this stage from the informal testing in the *fourth* stage in that the main aim of the test is to provide evidence which can be used to guide decisions about the future work of the pupil. For a test at the end of one of the sequence of years within a school, such decisions might be about whether or not to continue the study of a particular subject in the next school year, or the choice in assigning each pupil to the most suitable class group in the next year. For a test at the end of the top level in that school, such decisions might be about the best choice of the next school or of future career after education. In these stages, and depending on the state system of education, the tests may be externally set and marked. There are extra dimensions involved here, as aggregated test scores may be used to appraise the work of a particular teacher, or of the school as a whole. So this stage is characterised by decisions and by high- stakes consequences.

The above outline may be summarised as presenting a model of pedagogy in terms of five stages as follows:

a. Decide learning aims, both for subject concepts and for the development of the learner.
b. Select and plan activities to reflect the aims and engage pupils' interest.

c. Implement in classroom interactive dialogue.
d. Review a topic or stage to consolidate and check for gaps using informal summative assessments.
e. Guide decisions about the next stages using formal summative assessment, perhaps high-stakes.

This model was composed to meet the need for a model of pedagogy in which the role of assessment could be made clear. I have argued elsewhere that some of the well-established works about the nature of pedagogy do not treat the role of assessment seriously, i.e. it is seen as a marginal component rather than a central one (Black, 2013). By contrast, the role(s) of assessment stand out clearly in the above. Formative feedback is a main component of stage C and the planning in stage B has to foresee how this can be ensured. The role in stage D might be a dual one insofar as test results will be used formatively to complete the gap-filling function, yet summatively to guide decisions. The role in stage E is clear, but the extent to which this stage is in the control of teachers and schools will vary between school years and between different state systems.

These five stages do not function as a linear sequence in the decisions that teachers make about their pedagogy. An activity designed in stage B might encounter problems on implementation in stage C which might lead to a new design in B for use next time. A deeper problem may arise if, on reflection, what seemed like a successful classroom activity had not helped pupils to think more carefully about the principles, e.g. of control of variables in an experimental investigation. This might be an example of imperfect match of the stage B planning to the stage A aims. In a similar manner, results of tests in stages D or E ought to be scrutinised to see whether they really provide evidence of achievement of the aims of stage A. One common problem here is that it is easy to specify very attractive, and often very vague and general, aims in stage A, but when these are compared with the actual evidence of the final test in stage E, it is evident that the latter only calls for a far more limited set of achievements. Indeed, some have argued for a 'bottom-up' approach (Klenowski & Wyatt-Smith, 2014). They reported the result of using this approach with teachers in the following way

> 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. 105)

Thus, any assessment instruments planned for stage E ought to be formulated at the first stage of curriculum planning, to ensure that there a match between these and the aims, or that any mis-match is explored and resolved before going further.

TEACHERS AS ASSESSORS

The view developed in the previous section is very different from the view of assessment held by many teachers. The common view is that assessment is the unpleasant dimension of learning, that there are stages when one has to stop helping pupils to engage in and enjoy their learning and start instead to 'teach to the test' with the implication that such teaching involves rote learning designed to anticipate the types of question which are often asked. This negative view is obviously most strong at times when external tests, set by bodies external to the school, are approaching.

Another cause of the negative views of testing held by many teachers is that study of the concepts and of the different possible methods of assessment is often neglected in teacher training. This is a serious fault because it means that teachers lack both confidence and skill in designing their own summative assessments. Yet for many years of schooling, years in which externally set tests do not operate, the quality of the summative assessments set and marked by teachers for use in their own classes is important in that the results are used to guide pupils – and to assess or guide teachers themselves.

A limited intervention study, lasting for two years, which aimed to help about 15 teachers, of mathematics or of English, from three schools, to survey and improve their own summative assessments, gave evidence of these difficulties (Black et al., 2010, 2011, 2013). Where externally set tests were not imposed, many teachers used sets of questions from past national tests, or from testing companies, for their end-of-year tests, without exploring the quality, or relevance to their teaching, of these instruments. What was needed was to focus their attention on the concept of validity, stressing that the main and over-arching criterion was that a high test mark would justifiably imply, to those who used the results, the degree to which each pupil was good at doing the subject. This debate was found to be valuable, as one teacher of mathematics reflected:

> It all points towards the 'what does it mean to be good at maths' question and how we [get] the students to show this – surely tests in a formal way (if properly constructed) have a role to play in allowing students to demonstrate this – and does also leave scope for teacher assessment – if the teachers are confident in this. (Black et al., 2010, p. 223)

The issue was whether doing well in the assessment was valid evidence of capability in doing the subject, so the teachers had to debate what being good at the subject meant to them. The consensus that they achieved, through their discussions of this topic, led them to realise that their summative assessments were not fit for purpose and needed to be re-designed. This was an example of fruitful interaction between stages A and E outlined in the previous section.

PUPILS AS LEARNERS

The main aim of this chapter has been to emphasise the several ways in which assessment helps pupils to develop as effective learners. Thus:

> *Feedback in oral dialogue* helps pupils' to engage, through discussion, in reasoning about their own ideas and about those of fellow-pupils**.**
>
> *Feedback in written dialogue* helps all pupils to reflect on their work and to think critically about the quality of their own arguments.
>
> *Feedback can promote confidence in learning* if a focus on specific comments on their work helps pupils to see feedback, not as a judgment on their 'innate ability', but as guide to improvement: this can encourage them to believe that they can improve performance by their own thoughtful efforts.
>
> *Peer-assessment by pupils* can contribute to their development of self-assessment, and thereby help them become more mature and independent learners, able to structure their own learning and reasoning.
>
> *Effective group work* enhances peer-assessment in its development of pupils 'mental self-regulation', whilst also making them more capable of contributing to, and learning from, group work. Finally,
>
> *Formative use of summative assessments* can help pupils to appreciate the value of an overall review of their work – by identifying faults in their understanding of that work and in correcting these before moving to a new topic.

One overall feature of this list is the stress on pupils being helped to take responsibility for, their own learning. The US author, Thomas Groome, has emphasised one dimension of this argument in the following way

> 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)

One of the bad effects of accountability measures, where these are implemented with formal testing which rewards a limited range of learning behaviours, is that they motivate teachers to 'teach to the test'. Such teaching can make their pupils, as Groome says, too dependent on the teacher, even although there is evidence that teaching with a broader focus on 'empowering learners' can produce better performance, even in such tests (e.g. in mathematics see Boaler, 2002).

Related, but more fundamental, issues are argued in the following extract from Stanley et al. (2009)

> …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 behaviors that are required in dealing with authentic assessment tasks. (p. 31)

This highlights two issues. Science education can make a strong contribution in helping pupils, through 'Authentic assessment tasks', to be more capable in dealing with a wide range of academic and work-place problems. To make this contribution, science teachers should aim to help pupils understand how scientists work, which means that they should engage them in open-ended inquiry. Such inquiry involves setting pupils a problem about a phenomenon which they have observed, or can be guided to observe, and asking them to find out what they can about the factors which give rise to, and/or about any interventions which can change, what is observed. Where the example is well chosen, and the pupils are given a minimum of guidance, they have to exercise both reasoned judgment, initiative and skills in collaboration, to make progress, and so to achieve the aim 'to integrate the many elements of performance behaviors'. This potential contribution of science education to the broader development of young people has been spelt out in detail by work within the European Community and the innovations that they have supported have been shown to achieve the aims set out in the quotation above (see, for example, Fibonacci, 2010).

A second issue is that formal and externally set tests cannot support or validly assess such activities, that work has to be done by the teachers themselves. It follows both that teachers need more help in developing their skills and confidence in their own assessments, and that national systems must support and endorse their judgments.

CHANGING ASSESSMENT: PRINCIPLES AND PERSONAL CHANGE

In 1998, the Swiss researcher, Perrenoud, wrote a profound critique of the 1998 Black and Wiliam review: his main point may be illustrated by the following extract:

> This [feedback] no longer seems to me, however, to be the central 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)

The point of this critique, the need to locate formative assessment in a broader view of pedagogy, has been addressed, both at greater length elsewhere (Black & Wiliam, 2009), but also in this chapter in the discussion of a model of pedagogy.

A different aim in this chapter has been to develop ways, that assessment can enhance learning, which are directly applicable in practice. These have implications for the day-to-day work of both teachers and pupils. Many teachers who have attempted to develop dialogue with and between pupils in their classrooms have said 'It's pretty scary': the way they perceived and implemented their role as teacher was changing, and there was fear that they might be losing control. However, teachers will only discover how such development can empower both themselves and their pupils if they are willing to hand over to their pupils more of the control of the work. As one teacher expressed it:

> I was focusing on the girls' understanding and not on their behaviour. I often found that once the understanding was there, the behaviour followed. (Black et al., 2003, p. 96)

Many have also found that pupils themselves have, at first, resisted the change in their roles that the new emphases on oral and written dialogue require, so that any changes in practice required sensitive handling. As Perrenoud put it:

> … a number of pupils … are content to 'get by'… Every teacher who wants to practice formative assessment *must reconstruct the habits acquired by his pupils*. (Perrenound, 1991, p. 92, Emphasis in the original)

Thus, whilst this chapter is in part a theoretical analysis, it is also a challenge to practice.

NOTE

[1] Throughout this chapter, the terms 'pupils' and 'students' will be regarded as equivalent. The former term will be used in the main text, but where the latter was used in the author's original, it will be retained in any quotation.

FURTHER READING

Alexander, R. (2008). *Towards dialogic thinking: Rethinking classroom talk* (4th ed.). York, UK: Dialogos.

Black, P. (2016). The role of assessment in pedagogy – and why validity matters. In D. Wyse, L. Hayward, & J. Pandya (Eds.), *The Sage handbook of curriculum, pedagogy and assessment* (Ch. 45; pp. 725–739). London, UK: Sage.

Black, P., & Atkin, M. (2014) The central role of assessment in pedagogy. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research on science education: Volume II* (Chapter. 38, pp. 775–790). Abingdon, UK: Routledge.

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: GL Assessment.

Klenowski, V., & Wyatt-Smith, C. (2014). *Assessment for education: Standards, judgment and moderation.* London: Sage.

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.

Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree.

REFERENCES

Alexander, R. (2008). *Towards dialogic thinking: Rethinking classroom talk* (4th ed.). York, England: Dialogos.

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

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: Springer.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, *5*(1), 7–74.

Black, P., & Wiliam, D. (1998b). *Inside the Black Box.* London: GL Assessment.

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

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education, 17*(2), 215–232. ISSN 0969-594X

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.

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.

Boaler, J. (2002). *Experiencing school mathematics: Traditional and reform approaches to teaching and their impact on student learning* (Revised and Expanded ed.). Mahwah, NJ: Lawrence Erlbaum Association.

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.

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

Fibonacci Consortium. (2010). *The Fibonacci Project: Disseminating inquiry-based science and mathematics education in Europe*. Retrieved French Academy of sciences [website] January 20, 2011, from http://www.fibonacci-project.eu

Groome, T. H. (2005). *Educating for life.* New York, NY: Crossroad.

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 and Practice,* (1), 85–102.

Ruiz-Primo, M. A., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment, 11*(3&4), 205–235.

Stanley, G., MacCann, R., Gardner, J., Reynolds, L., & Wild, I. (2009). *Review of teacher assessment: What works best and issues for development.* Oxford, England: Oxford University Centre for Educational Development; Report commissioned by the QCA

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