# Using Assessment Materials to Stimulate Improvements in Teaching and Learning

**Robin Millar**

## Introduction

It is widely accepted that the quality of school science education depends on three interrelated elements: the curriculum (what we aim to teach), pedagogy (how we teach) and assessment (how we evaluate what students have learned). Major curriculum development projects in many countries have tended to focus on the first two. The third element, assessment, is often considered after the curriculum content, teaching approaches and materials have been developed. This can significantly undermine the overall success of an innovative development (for fuller discussion of an example of this, see Millar 2013). Several major innovations at national policy level in UK school science over the past few decades have failed to meet their designers' expectations because insufficient attention was given to the issue of how the intended student learning might be assessed. Two examples are the introduction of investigative practical work (Donnelly et al. 1996) and of a strand on 'how science works' in curricula for the 14–16 and 16–18 age range (Hunt 2010). This chapter discusses a current development and research (D&R) project in England (*York Science*) for lower secondary school (students aged 11–14) which starts from assessment.

R. Millar (✉)
University of York, York, UK
e-mail: robin.millar@york.ac.uk

## Context

Why did the *York Science* project choose to focus on the 11–14 age range? The reasons lie in the importance of this phase of students' science education in shaping their views and aspirations (Archer et al. 2013) and in providing the foundation for their future science learning. Successive changes to the curriculum framework in England for this age group, however, are widely seen as having led to a loss of coherence and clarity of purpose about learning goals (Oates 2010). The curriculum reforms initiated by the incoming government in 2010 sought to address this, by emphasising teaching and learning of the core ideas of science and raising attainment targets to match those of other leading jurisdictions worldwide. Consultations around the proposed changes were protracted, and the new national curriculum for students aged 11–16 and the associated assessment framework were not finalised and implemented until September 2016. A major concern for teachers is how to modify their programmes to address these changes.

## Theoretical Rationale

Although assessment is often associated with tests and examinations, its role in education is much broader. Kellaghan and Greaney comment that 'The term "assessment" may be used in education to refer to any procedure or activity that is designed to collect information about the knowledge, attitudes or skills of a learner or group of learners' (Kellaghan and Greaney 2001: 19). Assessment is a crucial aspect of the educational process because there is a very large difference between what is taught and what is learned. As Wiliam (2010) puts it:

> If what students learned as a result of the instructional practices of teachers were predictable, then all forms of assessment would be unnecessary; student achievement could be determined simply by inventorying their educational experiences. However, because what is learned by students is not related in any simple way to what they have been taught, assessment is a central – perhaps even *the* central – process in education. (Wiliam 2010: 254)

In discussions of assessment, three main purposes are often distinguished. Assessment may be used for summative purposes (to measure students' attainment at a specific moment – e.g. the end of a year, or term, or course – in a form that can be reported to the student and to others), or for formative purposes (to collect evidence of students' learning and use it to guide and encourage the subsequent actions of students and teachers), or for accountability purposes (to provide evidence of the effectiveness of teachers, schools and education systems). But lying behind all of these is a more fundamental role of assessment: to clarify the intended learning objectives of a teaching episode. Clarification is necessary because much of what is said and written about intended learning is ambiguous and unclear. For example, Mulhall et al. (2001: 583) ask: 'What, in detail, do we expect students to learn when

we talk of "conceptual understanding" in electricity'? They go on to argue that 'we do not have even the beginnings of systemic answers' but that 'some justified response to [this question] is a necessary, if not sufficient, condition for any helpful advances in the thinking about and practice of teaching electricity' (ibid.). This does not apply only to teaching and learning about electricity. The same could be said about any science topic.

Assessment is the tool that clarifies learning objectives; 'by its very nature assessment *reduces ambiguity*' (Wiliam 2010: 254, emphasis in original). A question or a task that we would expect students to be able to accomplish after instruction, if learning has been successful, provides the clearest indication of what the learning objective really means.

In addition to the key role of assessment in clarifying objectives, there is also a considerable body of research evidence showing that the formative use of assessment by teachers is associated with significant gains in student attainment (Black and Wiliam 1998; Hattie 2009). The impact of formative assessment on learning outcomes, however, depends crucially on how well the assessment is embedded in classroom practice and on the quality of the questions asked (Wiliam 2011). Wiliam concludes that 'sharing high quality questions may be the most significant thing we can do to improve the quality of student learning' (Wiliam 2011: 104).

Because assessment tasks provide the greatest clarity about learning objectives, Wiggins and McTighe (2006) advocate a 'backward design' approach to the planning of instruction. They argue that the first step in the development process is to write the questions or tasks that students should be able to tackle successfully at the end of a teaching episode and only then begin to think about how to teach to get them there. From a curriculum developer's perspective, specifying exactly how the intended learning outcomes of a course or module will be assessed is the best way to make clear to potential users what these outcomes are and mean. This then enables more focused and effective teaching and in the longer run enables a more focused evaluation of the effectiveness of the approach that underlies the development and of the materials developed to help implement it.

These lines of thinking provide the rationale for a curriculum project that centred on the development of a large, structured set of diagnostic questions and tasks as a resource that might both facilitate and, at the same time, shape teachers' classroom actions and their longer-term planning.

## Research Question and Methods

The central research question which the project addressed was:

- In what ways are teachers' practices and views changed by providing access to structured banks of diagnostic assessment resources?
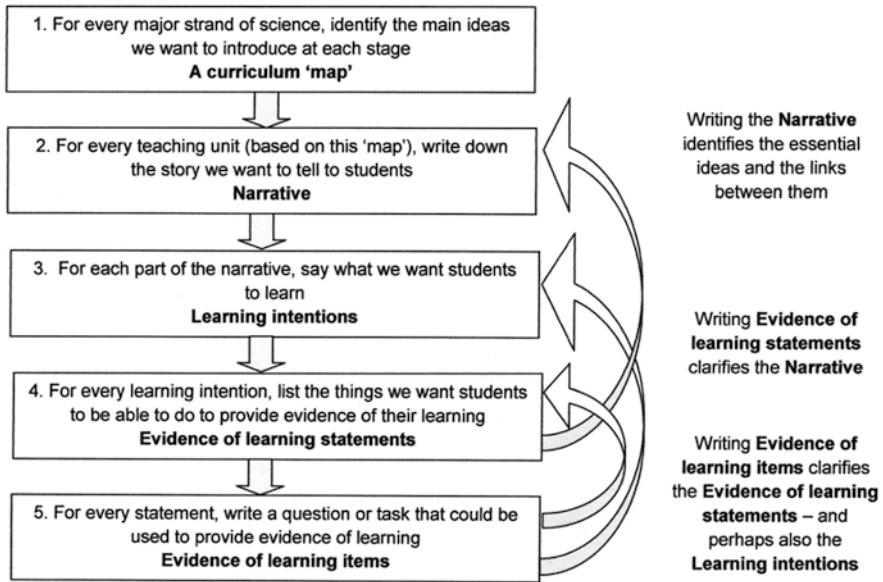
**Fig. 1** The development process used in the York Science project

## Development Phase

The development process which the *York Science* project adopted is shown in Fig. 1. As the project was dealing with a 3-year period within a 5–16 continuum, the first step was to develop a curriculum 'map' outlining how the major ideas in each of the main strands of science content might be expected to develop over the 5–16 age range. A 'main strand of science' here means a major topic like forces and motion, electricity and magnetism, chemical change or evolution. This 'map' in effect proposes an outline teaching sequence or learning progression (Corcoran et al. 2009).

The 'map' developed for *York Science* was influenced (though not totally constrained) by the requirements of the national curriculum but also informed by the available research evidence on students' learning (AAAS 2001; Driver et al. 1994; Duit 2009; Victoria State Government 2014) and by professional experience. A teaching sequence for the whole 5–16 age range enabled principled decisions to be made about the ideas to be introduced and developed in each strand within the 11–14 age range, which was the project's target, and made explicit what we assumed would have been taught by age 11 and what should be left until after age 14.

The second step was then, for each strand of science, to write down the story we want to tell to students at the 11–14 stage, as a continuous *narrative*. This is much more useful than a list of learning targets or objectives. Setting out the story briefly, but clearly, helps to identify the main ideas that have to be included and to sort out the order in which they need logically to come and the links between them. Although *narratives* were written with teachers (not students) in mind as the audience, they

Some objects can affect others at a distance by emitting *radiation* which travels from one object (the *source*) to another (the *receiver*), through the material or the space (the *medium*) between them. Light and sound are examples of radiation. Radiation travels out from a source in straight lines in all directions. When it strikes another object, it may go straight through (*transmission*), bounce off (*scattering* or *reflection*), or be stopped (*absorption*) – or a combination of these. When radiation is blocked by an *opaque* object, this causes a shadow region. The effects of radiation get steadily less the further it goes, because it is spread over an ever-increasing area, and because it may be gradually absorbed by the medium it is travelling through. When radiation is absorbed by an object, it has an effect on the object; this might be a chemical effect, an electrical effect or a heating effect.

**Fig. 2** The first part of the York Science Narrative on Radiation (Light and Sound)

**Table 1** Sample evidence of learning statements for the topic Radiation (Light and Sound)

| Evidence of learning statements |
| --- |
| *Students should be able to:* |
| Identify, in a given situation, the radiation source, the receiver of radiation and the medium through which radiation is travelling |
| Explain and predict the shapes and sizes of shadows cast by point sources |
| Explain why a source of light becomes dimmer (or the sound from a source fainter) the further you are away from it |
| Use the radiation model to explain familiar phenomena and events and to make predictions |
| Identify evidence for specific aspects of the radiation model |
| Offer plausible suggestions about what will happen in a given situation when radiation is absorbed: (i) to the absorber, (ii) to the radiation |

use the language that would be used in 'telling the story' to students. An illustrative example of part of a *narrative* is shown in Fig. 2.

A *narrative* usually consists of a sequence of paragraphs (or sections). For each section, the next step (step 3 in Fig. 1) is to say briefly what the learning intention for that part of the story is: what we want students to learn. For the *narrative* section in Fig. 2, this might be that 'Students should understand and be able to use the source-radiation-receiver model'. This is then followed by two crucial steps. First (step 4 in Fig. 1), the learning intention is translated into a set of observable performances: a list of things we would expect students to be able to do if their learning has been successful. This step, in effect, involves operationalising the learning intentions. Words like 'know' and 'understand' disappear and are replaced by the observable actions that we would take as evidence of knowledge and understanding. To illustrate this, some *evidence of learning statements* for the *narrative* section shown in Fig. 2 are listed in Table 1.

Finally, and equally crucially, step 5 (in Fig. 1) is to write at least one question or task that a teacher could use in class to obtain reasonably good evidence of students'

learning, quickly enough to be able to use this to inform the next actions of the students and/or the teacher. We called these *evidence of learning items*. Among the formats used were:

- Two-tier multiple-choice questions
- 'Talking heads' questions, where students are asked to evaluate a set of responses to a situation presented in speech bubbles and in terms that a student might use
- Predict-explain-observe-explain practical tasks
- Confidence grids: questions in which several statements are made about a given situation and students have to put each statement in one of the categories (I'm sure this is right/I think this is right/I think this is wrong/I'm sure this is wrong)
- Construct an explanation: where students have to select the correct option in each of a sequence of boxes to construct a correct explanation of a given event or phenomenon
- Evaluating a representation: where students have to identify aspects of a given representation (usually a textbook diagram) which they think are 'a good picture' of the real thing and aspects which they think are not

This list is not complete; *evidence of learning items* of other types and formats have also been developed.

As the right-hand side of Fig. 1 emphasises, this is an iterative process, not a linear one. Writing *evidence of learning items* often makes you question the way the corresponding *evidence of learning statement* has been expressed or helps you notice that a statement is missing and should be added. In some cases, this indicates a need to revise the stated *learning intention*, or even the *narrative*. The outcome of the development process is a large set of *evidence of learning items* for each of the main strands of science, linked clearly to (and consistent with) a *narrative*, a set of *learning intentions* and a list of *evidence of learning statements*.

At the time of writing, this development work has been completed for around half of the biology, chemistry and physics content required by the English national curriculum for the 11–14 age range. Work is continuing on the remaining strands.

## Research Phase

To obtain evidence of the impacts of the project's approach and materials (and to obtain feedback to improve these), teachers in 45 schools were given a large set of *evidence of learning items* (ELIs) for one of the first three science strands developed. The three sets were allocated randomly to schools. They were accompanied by guidance material which encouraged teachers to use the ELIs for formative purposes, rather than summative ones, and suggested a range of ways of using ELIs that preliminary work had shown to be valuable. In particular, teachers were encouraged to use *evidence of learning items* as stimuli for small-group discussion rather than as individual written exercises or tests and to see the discussion these generated as a valuable source of evidence of students' thinking.

After the teachers had had the material for around 3 months, written question-naires (n = 45), augmented with interviews where this was feasible on grounds of availability and access (n = 13), were used to collect feedback including:

Descriptive data (What had they used? How had they used it?)
Evaluative data (What did they think of the materials? Any suggestions for improve-ment/addition?)
Reflective data (How might this change their teaching and planning?)

Responses were analysed, initially using predetermined categories implicit in the data collection instruments, modified by an inductive analysis using a grounded theory approach to pick up any unanticipated themes (see, e.g. Robson 2002).

## Findings

The reception by teachers of the project materials and approach has been strongly and uniformly positive. Almost all saw the project as directly relevant to issues with which teachers are currently grappling as a result of policy-driven changes.

Many teachers said they were aware of common 'misconceptions' (the term they invariably chose to use) that some students are likely to hold, but several expressed surprise at their prevalence. One commented that 'without the questions, I might never have been aware how widespread particular misconceptions were' (T09). Others expressed surprise that many students did not understand things they expected them to have grasped. One wrote:

> When I was given the trial pack to try it out, I was in the middle of teaching light and I thought "Oh, I'll try some of these, they'll be able to do them, no problem for students." But they couldn't. (T02)

She followed this up by sending the response of a student group to a question designed to probe ideas about primary sources of light. It asked what you would see if you closed yourself inside a dark cupboard with a well-sealed door and no win-dow, in which there was a mirror and a cat. Students were given four statements to evaluate. Writing of her class, she said: 'One student got it right, the most common response by far was this' (Table 2).

Reflecting on this and other items on the same topic, this teacher commented that:

> I really like how I'm able to get down to the nitty-gritty of what the kids are thinking … how are they actually thinking about it? There's an activity about light travelling in straight lines and where it travels from, and they all thought that light comes out of your eye. I really thought that they would know all of this, there'd be no problem with the science, and oh my goodness there were problems with the science. That was really eye opening and I really liked that. I thought, if this can tell me about things that I thought students would know then what could it tell me about the things that I'm actually teaching them? (T02)

Other respondents also replied that using items from the question bank not only showed them what many students thought but also gave them insights into the thinking

**Table 2** Data from a teacher on the commonest response pattern in her class to one evidence of learning item

| | Statement | I am sure this is right | I think this is right | I think this is wrong | I am sure this is wrong |
|---|---|---|---|---|---|
| 1 | After a whilst, you will be able to see everything, but very dim | | ✓ | | |
| 2 | The only thing you will see is the cat's eyes shining | | ✓ | | |
| 3 | You will see the mirror shining, but everything else will be dark | | | ✓ | |
| 4 | You won't be able to see anything at all, no matter how long you wait | | | | ✓ |

that lay behind their answers. As one teacher put it, 'It makes you look at things from an understanding level and also informs you on an understanding level as well' (T01). He commented on some benefits of research-informed multiple-choice questions:

> The nice thing about this, it's multiple choice, you have various different answers, but there are some which if your thinking isn't quite right, that's the one you'll go for. And that's really really helpful and really useful. You can listen to the thought processes, they have discussions about it, what do you think, what's this, how does this work, and that really helps you into what they're thinking and how it works.' (T01)

This teacher went on to talk about how his use of a set of items on chemical substances and chemical change were changing the way he taught this topic:

> The YS materials pick up the misconceptions in such a way that it's clear what they don't understand and how they don't understand it. So it's better than simply getting a wrong answer on a test, you've actually got some sort of idea about what they don't understand and a potential way in to fix it. And it's mainly go back to the particular lessons where I knew there was a problem and take another look at them as well. When I've taught it again, I've approached it in a different way. (T01)

Another point made by several teachers was about the value of these questions in stimulating well-focused discussions in student groups. One remarked that 'they [the questions] were so interesting to use. The use for me is opening up the discussion, thinking about how they're actually perceiving things, that was the interesting bit' (T11). Another commented that 'so much of what is generated from this is discussion with the pupils … It's prompted more discussions than I would normally have had … which is good' (T07).

Whilst there are many challenges in designing good diagnostic assessment questions and tasks, teacher feedback has not reported any significant challenges or problems in using the materials produced by the *York Science* project.

## Conclusions and Implications

This preliminary and small-scale evaluation of the *York Science* materials for three science strands, and of the embedded formative assessment approach that they promote, provides encouragement that carefully designed and research-informed assessment materials can have the intended impacts on teachers' practice and thinking. In general, the materials were used as intended, for assessment during lessons 'in real time' and as stimuli for 'on-task' discussion. This study provides 'proof of principle' that the strategy the project has adopted that can work indeed is quite likely to work. This strategy might be summarised as seeking to stimulate changes of specific and planned kinds in teachers' practices, by providing resources which make it easier for them to implement these changes and hence to influence their thinking about teaching and learning more generally and about the planning of lesson sequences.

The responses of teachers to the project materials and approach confirm the view from which we began that assessment items play a crucial role in communicating intended learning objectives clearly and that structured sets of items are particularly valuable in focusing teachers' attention on learning outcomes and facilitating the use of embedded formative assessment to monitor students' ideas and learning as the teaching proceeds and respond to this evidence 'in real time'.

## References

AAAS (American Association for the Advancement of Science). (2001). *Atlas of science literacy*. Washington, DC: AAAS.

Archer, L., Osborne, J., DeWitt, J., Dillon, J., & Wong, B. (2013). *ASPIRES. Young people's science and career aspirations age 11–14* (Final report). London: King's College London.

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

Corcoran, T., Mosher, F., & Rogat, A. (2009). *Learning progressions in science: An evidence-based approach to reform* (CPRE research report#RR-63). Philadelphia: Consortium for Policy Research in Education.

Donnelly, J., Buchan, A., Jenkins, E., Laws, P., & Welford, G. (1996). *Investigations by Order. Policy, curriculum and science teachers' work under the Education Reform Act*. Nafferton: Studies in Education Ltd.

Driver, R., Rushworth, P., Squires, A., & Wood-Robinson, V. (1994). *Children's ideas in science*. London: Routledge.

Duit, R. (2009). *Bibliography STCSE. Students' and Teachers' conceptions and science education*. Kiel: IPN. Retrieved June 27, 2016 from archiv.ipn.uni-kiel.de/stcse/

Hattie, J. (2009). *Visible learning*. London: Routledge.

Hunt, A. (2010). *Ideas and evidence in science: Lessons from Assessment.* Report prepared for SCORE (Science Community Representing Education). Retrieved April 13, 2016 from www.score-education.org/media/7376/finalhsw.pdf

Kellaghan, T., & Greaney, V. (2001). *Using assessment to improve the quality of education*. Paris: UNESCO. International Institute for Educational Planning.

Millar, R. (2013). Improving science education: Why assessment matters. In D. Corrigan, R. Gunstone, & A. Jones (Eds.), *Valuing assessment in science education: Pedagogy, curriculum, policy* (pp. 55–68). Dordrecht: Springer.

Mulhall, P., McKittrick, B., & Gunstone, R. (2001). A perspective on the resolution of confusions in the teaching of electricity. *Research in Science Education, 31*(4), 575–587.

Oates, T. (2010). *Could do better: Using international comparisons to refine the National Curriculum in England*. Cambridge: Cambridge Assessment.

Robson, C. (2002). *Real world research* (2nd ed. Ch. 14: The analysis of qualitative data, pp. 455–499). Oxford: Blackwell.

Victoria State Government (2014). Science Continuum F-10. Retrieved April 13, 2016 from www.education.vic.gov.au/school/teachers/teachingresources/discipline/science/continuum/Pages/default.aspx

Wiggins, G., & McTighe, J. (2006). *Understanding by design* (2nd ed.). Upper Saddle River: Pearson.

Wiliam, D. (2010). What counts as evidence of educational achievement? The role of constructs in the pursuit of equity in assessment. *Review of Research in Education, 34*, 254–284.

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