Issues in policy and practice in the assessment of inquiry-based science education

Wynne Harlen, OBE, PhD

INTRODUCTION

The increased attention to assessment of students' achievement in the past 30 or so years, although welcome in many respects, has thrown up very many questions, challenges and matters for research (the meaning of 'issues' in this context). In this paper I first briefly raise four dilemmas facing assessment in general, before turning to questions that rise in relation to assessment in inquiry-based science education (IBSE) in particular. The four issues arise from conflicts, or disconnects, between what we would like, or ought, to be able to accomplish through assessment and what happens in practice. They are:

1. The disconnect between the aims, goals and values of education and what is currently assessed.
2. The disconnect between the aspirations of using assessment formatively and the reality of the predominance of summative uses.
3. The disconnect between the way in which we assess and what we understand about how students learn.
4. The disconnect between the narrow range of goals currently assessed and the goals of education in a global context.

Issue 1: The disconnect between the aims, goals and values of education and what is currently assessed

Assessment was once regarded as something that takes place after learning and as being quite separate from the process of learning. This view is no longer tenable; assessment is now acknowledged as a central part of education, with a proven role in helping learning as well as in reporting it. How the results of student assessment are used is recognised as having an important influence, which can be positive or negative, on the content and methods of teaching. The now well-acknowledged relationship is indicated by the equally familiar triangle in Figure 1.

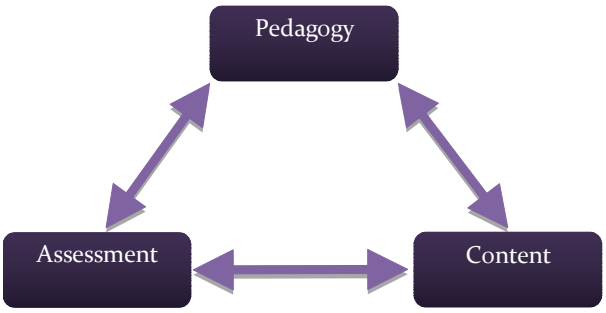

Figure 1: Interactions among curriculum content, pedagogy and assessment

The arrows acknowledge what is well known – that what we teach is influenced by how we teach, and what and how we assess influences both how and what we teach. These interactions are important for it is no use advocating the use of inquiry-based teaching if there is an overbearing assessment (whether by testing or other means) or a curriculum overcrowded with content. It is no use suggesting that the content should be focused on 'big' ideas if the assessment requires memorising multiple facts or if the pedagogy does not forge links that are necessary to form these big ideas; and it is no use wanting students to develop responsibility for their own continued learning if teaching does not allow time for reflection and room for creativity. Nor can we hope for positive attitudes towards science if the curriculum content seems to students to be remote from their interests and experience.

This does not mean that the impact of assessment on the curriculum content and teaching approach is necessarily a negative one. An effective assessment system supports learning in a variety of ways, from providing formative feedback for use in short-term decisions about learning activities to providing information about students' achievement for reporting to parents and others, for use in longer-term planning and as part of school self-evaluation. Furthermore, the process of assessment can help to clarify and communicate the meaning of learning objectives through establishing criteria for achievement or providing tasks that exemplify the use of inquiry skills and understanding.

But unfortunately negative impacts all too frequently arise. They generally result from assessment tools falling short of enabling students to show what they know and can do in relation to the learning goals. In the context of inquiry-based education it is a matter of concern that most current assessment tools and procedures fall short of what is needed to provide a good account of students' achievement of the goals of IBSE. The negative impact of this deficiency is compounded when the results for 'high stakes' evaluation of teachers and schools. When rewards and punishments are attached to assessment results this puts pressure on teachers, which is transferred to students, even when the results are not high stakes for students (as in sample surveys). Research shows that when this happens, teachers focus teaching on what is assessed, train students in how to pass tests and feel impelled to adopt teaching styles which do not match what is needed to develop real understanding. There is now a large body of research evidence on the negative impact of high stakes use of data from assessment and testing.

To engage effectively with this issue we need to develop assessment strategies and tools that better match the content and pedagogy of 21st century education. But more than this, it needs a change in policy to cease using student outcomes as the sole measure of quality of teaching or school provision for learning. One reason for this is simply that what students achieve is influenced by many factors as well as their school experiences. Another reason is that It does not provide evidence that is needed about students achievements or the quality of teaching. There has been plenty of evidence accumulated over the 25 years of experience of testing to show that year-on-year increases that appear in test scores immedately after immediately after introducing high stakes national testing are due to familiarity with test-taking and to teaching to the test (Tymms, 2004; Linn, 2000). Test scores may rise – at least at first – but this does not give information about change in real learning. The consequence of focusing on what is tested, practising test-taking and the restricted range of what is tested, is that it is not really possible to tell from national test results whether or not national standards have changed year-on-year. In other words the high stakes use of the measure defeats purpose of using it. Instead national sampling surveys, using a

wide range of assessment tools, as now practised in many countries, provides a far better picture of national performance.

Issue 2. The disconnect between the aspirations of using assessment formatively and the reality of the predominance of summative uses

WE HAVE TO CONSIDER THE TWO MAIN PURPOSES OF ASSESSMENT:

- to help students while they are learning
- to find out what they have learned at a particular time.

FORMATIVE ASSESSMENT

Formative assessment has the purpose of assisting learning and for that reason is also called 'assessment for learning' (AfL). It involves processes of 'seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning and where they need to go and how best to get there' (Assessment Reform Group, 2002).

What is involved in formative assessment can be described in terms of an on-going cyclic process (Figure 2) in which information is gathered in relation to the students' progress towards the short-term goals of a lesson or series of lessons. This information is then used to identify the appropriate next steps for the students and the action needed to take these steps. Students, of course, are the ones who do the learning so a key feature of formative assessment is the feedback that students receive about how to improve their understanding or skills or move on. At the same time the information gathered about students' progress provides feedback to the teacher, who can then adjust the pace or challenge of the learning activities – or regulate the teaching – to maximise opportunities for learning. Students, too, can have a role in decisions about their learning and direct their efforts more effectively if they know the purpose of their activities. This means not just knowing what to do but what they are trying to achieve in terms of quality as well as goals.

In summary, the key activities that formative assessment involves are

- Students being engaged in expressing and communicating their understandings and skills through classroom dialogue, initiated by open and person-centred questions
- Students understanding the goals of their work and having a grasp of what is good quality work
- Students being involved in self-assessment so that they take part in identifying what they need to do to improve or move forward
- Feedback to students that provides advice on how to improve or move forward and avoids making comparisons with other students
- Teachers using information about on-going learning to adjust teaching so that all students have opportunity to learn.
- Dialogue between teacher and students that encourages reflection on their learning

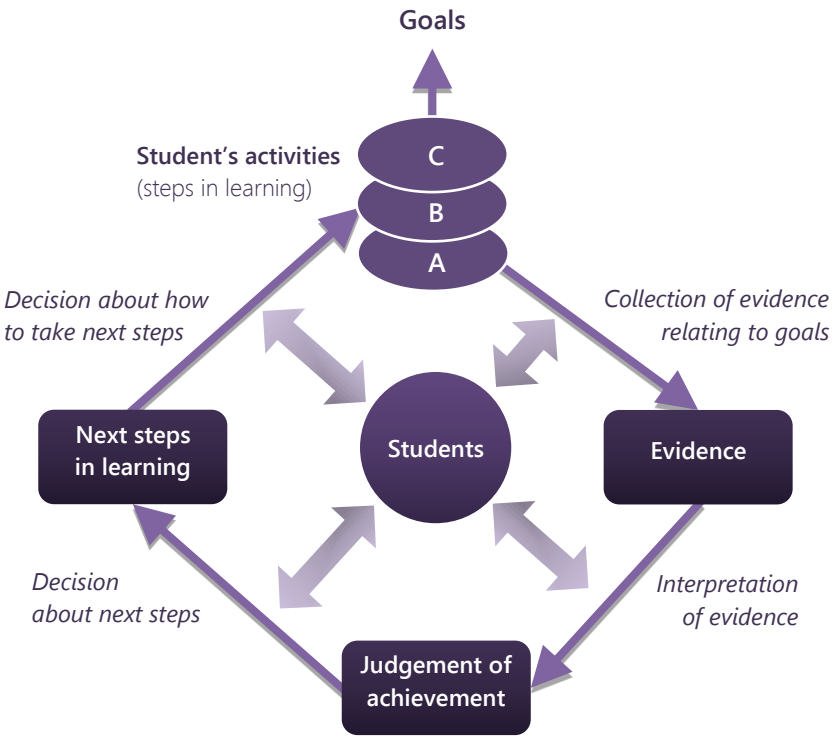

Figure 2: Assessment for formative purposes

The reason for attention to formative assessment lies in the evidence of its effectiveness in improving learning. Empirical studies of classroom assessment have been the subject of several research reviews. The review by Black and Wiliam (1998) attracted attention world-wide partly because of the attempt to quantify the impact of using formative assessment. A key finding was that 'improved formative assessment helps the (so-called) low attainers more than the rest, and so reduces the spread of attainment whilst raising it overall'. Since then there have been a number of other reviews and investigations which have justified the considerable claim made for improved student learning.

SUMMATIVE ASSESSMENT

Summative assessment has the purpose of summarising and reporting what has been learned at a particular time and for that reason is also called 'assessment of learning' (AoL). It involves processes of summing up by reviewing learning over a period of time, and/or checking-up by testing learning at a particular time.

Since formative assessment is defined as helping learning, there is tendency to regard it as the 'good' face of assessment and summative assessment as the reverse. But summative assessment is important for a number of reasons. First, whilst it is not intended to have direct impact on learning as it takes place, as does formative assessment, it nevertheless can be used to help learning in a less direct but necessary way as, for example, in providing a summary of students' learning to inform their next teacher when students move from one class or school to another. Second, it enables teachers, parents and schools to keep track of students' learning, both as individuals and as members of certain groups (such as those who are high achievers

and those who need special help). Third, it provides data which, together with contextual factors, can be used for school evaluation and improvement. Finally, it cannot be avoided: teachers have little choice about whether and when they conduct summative assessment since requirements and procedures are generally established at school or national level, not by individual teachers. By contrast, formative assessment could be considered, in a sense, to be voluntary in that it is possible to teach without it and it is part of the process of teaching, which teachers largely decide for themselves. Formative assessment can be urged in official documents but cannot be mandated in the way that summative assessment can be required by statute.

One reason for the poor reputation of summative assessment is that when measured performance becomes the dominant factor in the classroom it drives out formative assessment practice. Pollard *et al*. (2000) noted that the introduction of national tests in England in the 1990s and the requirement for teachers to assign levels to students affected their response to students and their use of formative assessment. Students were aware that whilst effort was encouraged, it was achievement on tests that counted. The same is found for older students right up to undergraduate level; student want to know ‘’is it for the examination?” to decide whether to give it their effort. Where there is competition between formative and summative assessment, the latter will always come out as the winner.

An obvious solution to this issue is to avoid competition by bringing the two together. The two ways of doing this are to make use of formative assessment data for summative purposes or to make formative use of summative assessment. Since formative assessment is carried out by teachers (and students), the first of these means using teachers’ judgments for summative assessment. In relation to the second case – of using summative assessment formatively, several ways of using classroom tests and internal school examinations to feed back into teaching and learning have been suggested (Black *et al.*, 2003). In practice the approach is one that teachers can use principally in the context of classroom tests over which they have complete control. Whilst some external tests and examinations can be used in this way, by obtaining marked scripts and discussing them with students, there is a danger that the process can move from developing understanding to ‘teaching to the test’ and in any case the feedback comes too late.

An example of combining formative and summative purpose in assessment that has high stakes for students is the approach used for many years in the Queensland Certificate of Education in Australia, used in determining entry to high education. In the Queensland system of externally moderated school-based assessment, teachers develop and implement assessment programs and instruments that cater for their school’s unique context, resources and students. The overall approach is the development of a portfolio of evidence from assessment tasks set by the teacher to meet the requirements of the syllabus for each subject. The portfolio allows a variation in the content so that syllabuses can be implemented with flexibility to meet local requirements. The common element is the system of progressive criteria, called Standard Descriptors, against which each portfolio is judged. The portfolio is built up over the two years of the course, during which time its content will change not only through addition of new material but through replacing older by more recent evidence. It is only the final evidence that is taken into account, although some will have been collected earlier than other. The criteria for assessment are published so that students and parents as well as teachers can be familiar with them. They describe what students can do in various categories and sub-categories at five levels or

standards. Evidence from the portfolio is compared with the criteria using 'on-balance' judgements of best fit.

Issue 3. The disconnect between the way in which we assess and what we understand about how students learn

The discussion in relation to Issue 2 focused mainly on the curriculum content. This issue relates more to pedagogy and the alignment between the contexts and processes of learning and of assessment. To explain why we need to bring learning theories into the discussion of assessment, consider the three main theories and their simple formulation (Watkins, 2003):

- Behaviourism: "Learning is being taught"
- Cognitive constructivism: "Learning is individual sense-making"
- Socio-cultural constructivism: "Learning is building knowledge as part of doing things with others."

Behaviourism describes a view of learning in which behaviours are formed by a system of rewards and punishments, so learning can be controlled externally and motivation is almost entirely extrinsic. A feature particularly relevant to assessment is that complex behaviours are deconstructed into parts which can be taught, practised and assessed separately. This view, then, is consistent with tests of disconnected facts and skills, where speed is of the essence and answers are either correct or incorrect.

Cognitive constructivism views learning as constructed by learners themselves and influenced by their existing knowledge. The aim is understanding, which is seen as occurring when new experience is incorporated into an existing or new model. The active participation of students is seen as paramount because, as widely quoted, 'they do the learning'. Constructivist views of learning underpin formative assessment, but there are few examples of summative assessment being based on a constructivist view of learning, although there are some attempts through computer adaptive testing and screen-based concept-mapping (Osmundson *et al.*, 1999).

In *socio-cultural constructivist* perspectives on learning there is also a focus on understanding but through 'making sense of new experience with others' rather than by working individually. In these situations the individual takes from (internalises) a shared experience what is needed to help his or her understanding, then externalises the result as an input into the group discussion. There is a constant to-ing and fro-ing from individual to group as knowledge is constructed communally through social interaction and dialogue. Modern views of science education reflect this approach, emphasising inquiry, thinking scientifically, building models, engaging in argumentation and critical reflection, through working in groups, sharing ideas communicating in a variety of modes. Clearly there is little in common between this view of learning and what is represented in traditional modes of assessment where students sit in isolation from one another in an examination room.

Some profound implications for assessment also follow from the view proposed by Vygotsky (1978) that for any learner there is an area just beyond current understanding where more advanced ideas can be used with help. Vygotsky called this area the 'zone of proximal (or potential) development'. It is, in essence, what we have called the 'next step' that the student can be expected to take identified through formative assessment. 'Scaffolding' is an apt term used to describe helping students to take this next step in understanding through introducing new ideas or better scientific

practices and providing vocabulary that enables students to express their ideas more precisely.

Recognising that, in the company of other learners, students can exceed what they can understand and do alone, throws into doubt what is their 'true' level of performance. Is it the level of 'independent performance' or the level of 'assisted performance' in the social context? It has been argued that the level of performance when responding to assistance and the new ways of thinking provided by others gives a better assessment than administering tests of unassisted performance (Grigorenko, 1998).

Research conducted in Denmark (Dolin and Krogh, 2010) using items from the PISA science tests provides clear support for this view. The research involved students in answering some PISA questions orally in an interview and conducting, in pairs, an investigation described in a PISA item. The conclusion reached was that 'when compared directly and following the scoring criteria of PISA, pupils' performance increased by 25% when they were allowed to exercise their knowledge in a socio-culturally oriented test format.'

Issue 4. The disconnect between the narrow range of goals currently assessed and the goals of education in a global context.

Issue 3 has taken the critique of assessment beyond concern with content. Issue 4 takes it further into matters that concern its contribution to goals of education that encompass major global issues: for example, the adverse impacts climate change and global warning on hunger, ill health, illiteracy, unemployment, etc. The question is: do we, can we, make any contribution to understanding and alleviating these conditions in the way we go about science education and its assessment? In one sense this seems a ridiculous question, like asking whether by eating less we can help the millions who go hungry across the world. But if we answer that we can do nothing, then what will ever change?

Education has a key role, particularly in developing students and future citizens who are thoughtful, in every sense, and understand the role of human activity in global warming, loss of diversity of organisms that lead to starvation, poverty, lack of education and unemployment. So we ought to be able through science education to make a contribution by helping understanding of the ideas that are relevant and powerful in making sense of the world and how it works, how its components interact, how human intervention can and cannot influence our global environment. This means identifying the 'big' ideas *of* science and *about* science (that is, how science operates, its strengths and limitations) (Harlen, 2010) and ensuring that science education is designed to develop understanding of these ideas.

So where does assessment come into the picture? In brief, it is through ensuring that all assessment helps learning. This means using assessment formatively to regulate teaching and learning to support understanding. It also means using summative assessment to support learning through better understanding of the goals and what it means to achieve them and monitoring the progress of students towards the powerful ideas and scientific inquiry skills. Assessment, then, needs to be part of the discussion of how to provide education of relevance to facing global problems.

CHALLENGES FOR IBSE

The goals of IBSE

These issues have wide relevance in educational assessment. To anchor the discussion to the context of inquiry-based science education (IBSE) we need to be sure of goals to be assessed. It is difficult to put together in a single statement the interacting components of the process of learning through inquiry, which is why there are several definitions. It is easier to show how understanding is built through collecting and using evidence to test possible explanations through a diagram, as in Figure 3.

> In inquiry-based learning the development of understanding stems from curiosity about a phenomenon or event (a) that is new to the learners and raises questions that grab their attention. Initial exploration may reveal features that bring to mind an idea from previous experience which suggests a possible explanation or an answer to a question (b). It may be the idea of an individual student or the result of brain-storming with other students or consulting sources of information. Working scientifically involves making a prediction based on the idea (c) and then gathering relevant data (d) to see if there is evidence to support the prediction and the application of the idea (e). This might be a lengthy investigation involving controlled experimentation or just a simple extension of observations.
>
> Finding that evidence fits with the prediction (f) and that the idea does provide a good explanation (b) means that this idea has become 'bigger' since it then explains a wider range of phenomena. Even if it does not seem to 'work', something has been learned about its range of application. But to find an explanation that does 'work' means that alternative ideas have to be used and tested. This may come from the initial or further brain-storming informed by what has then been found. The usefulness of the ideas developed in this way depends on the collection and use of evidence in a scientific manner. Thus the ability to use science inquiry skills is an essential part of the development of understanding and an outcome of shared thinking about what data to collect and how to go about collecting and interpreting them.

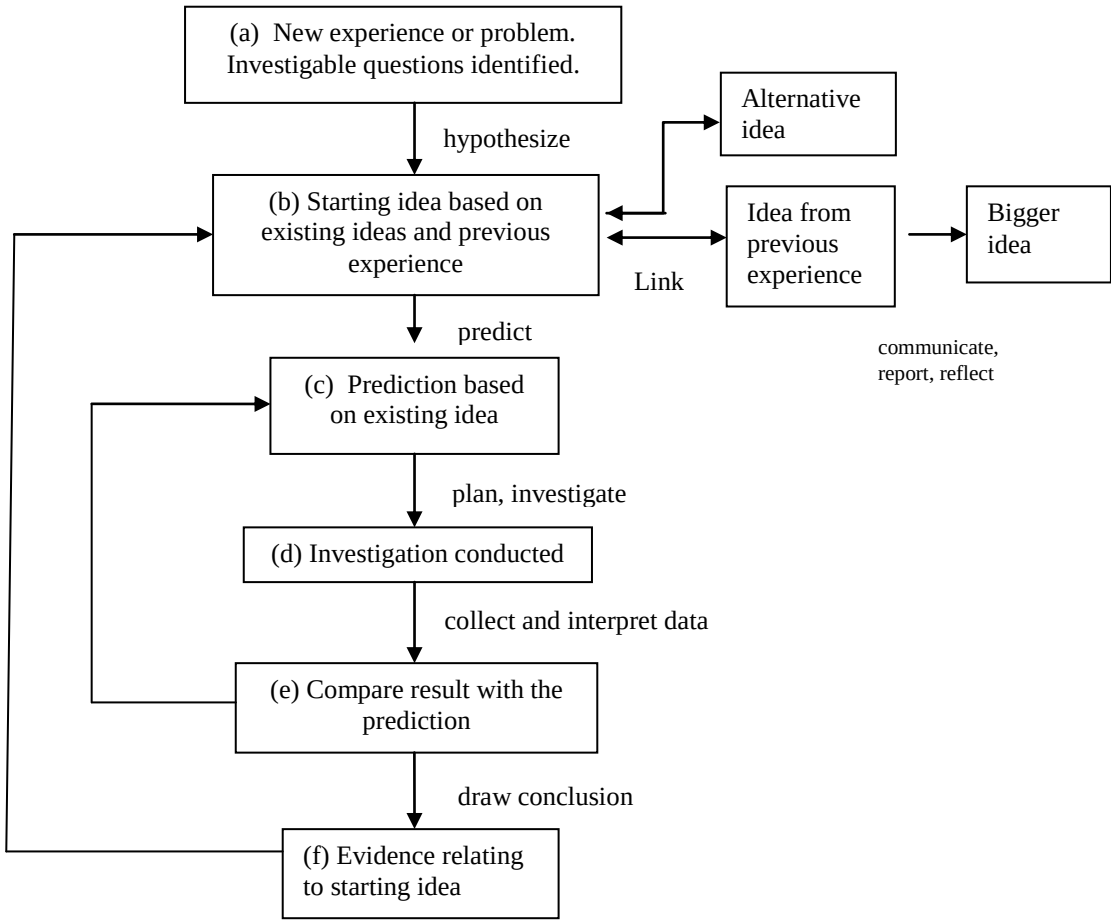

Figure 3: A model of learning through inquiry. Based on Harlen and Qualter (2013).

This description of inquiry does not restrict it to practical activity. Often the evidence that is needed will come from secondary sources rather than direct contact with or experimentation with materials. This challenges the assumption that inquiry must mean 'hands on' or 'practical work'. Another misconception, which is important in relation to assessment is that the aim of inquiry-based work is chiefly to develop the ability to 'behave as a scientist' and learn about a supposed 'scientific method'. There are two problems here. One concerns the goals of inquiry-based education in science. Placing the emphasis on processes of inquiry has led some to the mistaken view that inquiry is more appropriate in the primary school than in secondary education. Whilst it is important for students to know how scientific knowledge is created, their learning must help students at all levels to develop ideas that help them to understand in the world around, the ideas of science, as well as ideas about science. The other problem is the assumption of a single scientific method. In studying different aspects of the world, such as cosmology or ecology, scientists work in different ways. There is no single formula for scientific activity and certainly none that includes mathematics and science and thus no single approach to inquiry-based education.

IMPLICATIONS FOR ASSESSMENT OF IBSE

The discussion of formation assessment (figure 2) shows that it is essential to the implementation of IBSE. It involves skills and knowledge in accessing students' on-going learning through questioning, classroom dialogue and observation. But the greatest challenge is in using this information to decide, and then take, any action needed to help progress towards learning goals. This requires knowledge and understanding of development in students' conceptual and procedural learning. Many teachers need help with this part of formative assessment.

However, the influence that summative can have on formative assessment clearly means that giving attention to formative assessment alone would be likely to have little effect. Indeed the experience of introducing genuine formative assessment in countries where there exists a strong dependence on external high stakes tests, bears evidence to this. Thus if learning in science is to be improved through IBSE and the use of formative assessment, it is necessary also to ensure that the summative assessment is consistent with the learning aims of IBSE.

It is the dual nature of the goals of IBSE, the combination of conceptual understanding and skills that presents one of the greatest challenge to summative assessment. It means that both understanding and skills need to be assessed and raises the question of whether these can, and should, be assessed separately or in combination. Indeed it can be argued that it is not possible to assess understanding without some skills being used and vice versa. The assessment of understanding of a concept the assessment task should require an explanation of an event or interpretation of data or a prediction involving application of some concepts. Thus there are some skills (explaining, interpreting, predicting) that are also involved. For assessing skills, the task has to require the use one or more of the inquiry skills, such as predicting, planning, carrying out an investigation or interpreting given data. However, it is not possible to assess skills without involving some knowledge of the subject matter of its use. (Using trivial, non-scientific content raises the question of whether a skill is a science inquiry skill if it is not used in relation to science subject matter). Thus there will always be some aspects of understanding and skill required in all tasks. What determines whether a task is essentially assessing understanding or skill will be the level of demand on one or the other, and the credit given to different kinds of responses in scoring.

A further factor to be considered is that for valid assessment students need to be working on tasks where some aspects of inquiry are involved. There should also be some novelty in the task so that they are using their knowledge or skill and not simply recall of information, reasons or procedures that have been committed to memory. Genuine inquiry takes place when students seek to answer a question that is new to them and to which they do not already know the answer. But who is to judge what is 'new' for a particular student? Can the response of a student created in isolation from the normal context of learning and interaction with others really reflect their capability? These are questions which apply to any assessment conducted through tests or examinations but particularly to IBSE.

The alternatives to tests depend on the fact that the experiences that students need in order to develop desired skills, understanding and attitudes also provide opportunities for their progress to be assessed. The key factor is judgement by the teacher. Assessment by teachers can use evidence from regular activities supplemented, if necessary, by evidence from specially devised tasks introduced to provide

opportunities for students to use the skills and understanding to be assessed. Such approaches have to include effective quality management procedures that assure acceptable levels of reliability and consistency across schools. Key conditions for such an approach are time for teachers to take part in moderation to ensure dependability of the results and respect for teachers' professionalism. Time spent in this way, however, is a valuable form of professional development in assessment. Experience in Queensland, for example, is that 'The most powerful means for developing professional competence is assessment is the establishment of regular professional conversations among teachers about student performance (Maxwell, 2004).'

Other alternatives affording opportunities for more valid assessment of the outcomes of learning through inquiry may emerge from developments in screen-based assessment, but are yet at an early stage. Promising research into relevant assessment methods is being conducted by the ASSIST-ME (2014) project (an EU FP7 project) involving 10 partners in 8 countries.

References

Assessment Reform Group (ARG) (2002) Assessment for Learning: 10 Principles. www.assessment-reform-group.org

ASSIST-Me (2014). Available from: www.assistme.dk

Black, P. and Wiliam, D. (1998b) Inside the Black Box. London: School of Education, King's College London

Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2003). Assessment for Learning: Putting it into Practice. Maidenhead, England: Open University Press.

Dolin, J., & Krogh, L. B. (2010): The Relevance and Consequences of Pisa Science in a Danish Context. *International Journal of Science and Mathematics Education*, 8, 565-592.

Grigorenko, E. (1998) Mastering tools of the mind in school (trying out Vygotsky's ideas in classrooms), in Eds R. Sternberg and W. Wiliams Intelligence, Instruction and Assessment: Theory and Practice. Mahwah, NJ: Erlbaum.

Harlen, W. (Ed) Principles and Big Ideas of Science Education (2010) Available from www.ase.org.uk in English, from www.fondation-lamap.org in French, and from www.innovec.org.mx in Spanish.

Harlen, W. and Qualter, A. (2013) The Teaching of Science in Primary Schools. Sixth Edition. London: Routledge.

Linn, R. L. (2000) Assessments and accountability, *Educational Researcher*, 29 (2) 4-16

Maxwell, G. S. (2004)'Progressive assessment for learning and certification: some lessons from school-based assessment in Queensland'. Conference paper. March, Nadi, Fiji.

Osmundson, E., Chung, G., Herl, H., Klein D. (1999) Knowledge-mapping in the classroom: a tool for examining the development of students' conceptual understandings. Los Angeles, California: National Centre for Research on Evaluation and Student Testing, University of California.
www.cse.ucla.edu/Reports/TECH507.pdf

Pollard, A., Triggs, P., Broadfoot, P., Mcness, E. and Osborn, M. (2000) What pupils say: changing policy and practice in primary education (chapters 7 and 10). London: Continuum

Tymms, P. (2004) Are standards rising in English primary schools? *British Educational Research Journal*, 30 (4) 477-94

Vygotsky, L.S. (1978) Mind in Society: The Development of Higher Psychological Process. Cambridge, MA: Harvard University Process

Watkins, C. (2003) Learning: A Sense-Maker's Guide. London: Association of Teachers and Lecturers.