Assessing primary science learning: beyond paper and pencil assessment

SO Wing Mui Winnie

Department of Science
The Hong Kong Institute of Education

Email: wiso@ied.edu.hk

Received 26 July, 2004
Revised 31 August, 2004

Contents

Introduction

This article discusses the foundations for the assessment of primary science learning with a focus on the place of assessment in the curriculum, the association between pedagogy and assessment, and the role of formative and summative assessment. Alternative strategies are recommended for the assessment of science learning at primary level, and some important qualities of meaningful assessment are suggested.

According to Hughes and Wade (1996), teachers need to assess for two main reasons. First, teachers need to gain information on the progress of individual pupils so that appropriate activities may be organized in the classroom to enhance their learning. Second, teachers need to provide records of progress for parents that are based on the data that is collected through assessment. These records can also be passed on to other teachers so that the schemes of work of different teachers may match the ability of the pupils more appropriately.

Although such information can be obtained through formal assessment, it is far more likely to come from informal and continuous assessment during class time. A large amount of teacher-pupil interaction takes place in class, as pupils ask and answer questions, request help and assistance, hand in work for checking and marking, and make dialogues and discussions. Teachers, therefore, obtain a certain amount of information on the abilities and achievements of their pupils by this means, and also find out about the difficulties that pupils face in their learning.

The place of assessent in the curriculum

The place of assessment in the primary science curriculum needs to be addressed. Reinhartz and Beach (1997, p.320) considered "The heart of the assessment step is alignment with other aspects of the planning and implementation steps of the curriculum development progress? In the recent development of the General Studies curriculum that covers science learning for Hong Kong primary schools (Curriculum Development Council, 2000), teaching, learning, and assessment are placed together within the discussion of the curriculum framework. This clearly indicates the importance of appropriate assessment strategies at the curriculum development stage. Novak, Mintzes and Wandersee (1999) view assessment as the fifth essential commonplace in education, along with the teacher, the learner, the curriculum, and the social environment. And Hollins and Whitby (1998) see planning, teaching, and assessment as a continuous cycle (Figure 1).

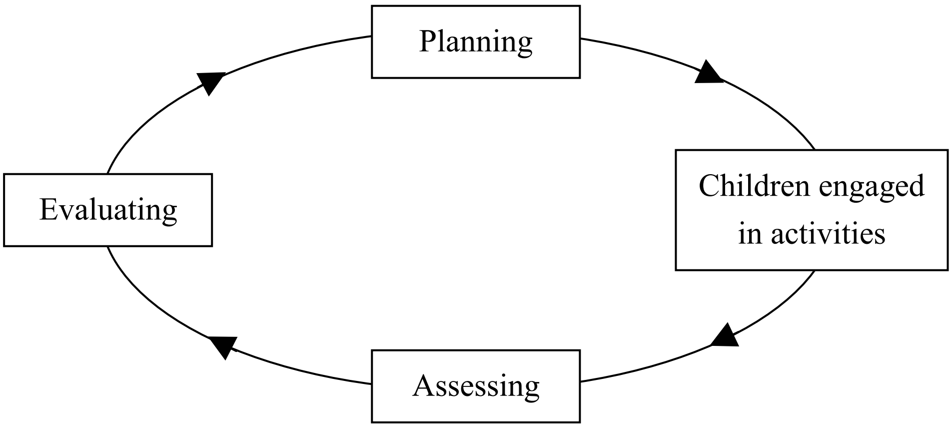

Figure 1: The assessment cycle (Source: Hollins & Whitby, 1998, p.153)

Association between assessment and pedagogy

Teaching and assessment are inseparable partners (Lowery, 2000). Teachers can often check on the progress of pupils as they walk around the classroom, and Hayes (1998) has pointed out that assessment and pedagogy in teaching and learning cannot be separated: "assessment should be at the heart of teaching and learning?(Hayes, 1998, p.125).

The objectives of most primary science curricula emphasize the understanding of science concepts, the use of process skills, and the application of knowledge to problem solving and science activities. Lee and Fradd (1998) suggest that there are three components of science learning: knowing science (scientific understanding), doing science (scientific inquiry), and talking science (scientific discourse). These components also include the integration of science learning with learning in other subjects at school. Black (2000) highlights that the practice of science education in the classroom requires attention to other curriculum disciplines, notably language skills and numeracy. Moreover, many science educators have advocated the implementation of the constructivist view of science learning. The personal constructivist view (Osborne & Wittrock, 1985; Glynn, Yeany & Britton, 1991) suggests that the learning of science is a process of construction and reconstruction of science concepts. From the social constructivist perspective, learning is a process of socialization or enculturation to shared knowledge through interaction (Bell & Gibert, 1996). No matter it is personal or social constructivist views, pupils' learning relies much on the continuous assessment of pupils' understanding during the learning process.

Unfortunately, Novak, Mintzes, and Wandersee (1999) claim that there has been a progressive decoupling, or "misalignment," of instruction and assessment in science education. In spite of this, there is an increasing understanding that assessment and teaching are interdependent (Reinhartz & Beach, 1997). As teaching and learning are strongly related to assessment, Black (2000) suggests that assessment policies should be formulated according to the curriculum-pedagogy-assessment triangle.

Formative and summative assessment

Planned assessment is often employed so that teachers can determine how well each pupil understands and how well pupils are doing. Lowery (2000) describes two common approaches. The first involves traditional evaluation strategies, which are usually known as summative assessment, that rely heavily on paper and pencil tests and examinations that rank the learning of pupils with scores and grades at a specific time in the school term. The second approach involves continuous and formative assessment strategies, with which teachers can check on the progress of pupils from time to time throughout the school year, and can thus obtain a general picture of what pupils understand and what they are able to do. Reinhartz and Beach (1997) suggest that assessment must help to improve teaching, and thus needs to be both formative and summative.

Summative assessment

Traditional strategies are commonly used to evaluate primary science achievement. These strategies include filling in the blanks for sentences and diagrams, matching components from different columns, judging items to be true or false, choosing the right answer from multiple-choice items, and giving short answers to questions, all of which are easy to administer and mark. The achievement of pupils in these traditional assessment strategies is only ranked by scores and grades, but Black (2000) comments that short, affordable, and externally set and marked tests cannot produce a reliable and valid assessment of a student’s capability except in particular and limited areas of science achievement. Lowery (2000) further states that these traditional strategies provide information about how well pupils recall knowledge and retain information, but do not allow for the expression of creativity or the development of original solutions to problems. It seems that the current practices in summative assessment are more deeply flawed than are generally realized with regard to both reliability and validity (Black, 2000). Moreover, the results from the traditional evaluation procedures give no information for either teachers or pupils on how to improve.

However, Hollins and Whitby (1998) argue that summative assessment involves a summing up of the point that a pupil has reached at the end of a particular time, such as at the end of a year or term. While Hayes (1998) opines that because of the emphasis on summative assessment in primary schools, planning and teaching with assessment in mind is developing slowly. There is thus a need to shift the emphasis to more meaningful assessment strategies to improve the association between assessment and pedagogy.

Formative assessment

The idea of formative assessment for science educators (Bell, 2000, Cowie & Bell, 1996, Black & Williams, 1998) is that the teacher gives feedback to the pupils, and then the teachers and pupils take action to improve learning during the learning process. Daws and Singh (1996) further elaborate that formative assessment is a process of learning by which pupils are encouraged to reflect on their learning in a structured and systematic fashion, and to discuss their progress with their teachers with a focus on what they need to do to improve.

Formative assessment is an inherent part of the teaching and learning process. It is usually informal and pupils are not aware of it (Hollins & Whitby, 1998). Information is gathered regularly in the course of day to day teaching, and is therefore used to make decisions about ongoing work for the pupils. It allows the teacher to adjust the activities that are given to the pupils to ensure an appropriate match. An illustration of formative assessment is given by Driver (1989), in which pupils are given feedback about how the concepts that pupils currently hold relate to scientifically accepted concepts, after which the teacher helps them to modify their thinking accordingly. This is both a part of formative assessment and of the teaching of conceptual development, and demonstrates the association between pedagogy and assessment.

In a review of studies of formative assessment, Black and Wiliam (1998) state that formative assessment does improve learning. However, Black (2000) finds that the use of formative assessment is still very weak, and ought to be developed further. In a review of literature on formative assessment, it is found that most assessment of this type focuses on individual authentic strategies of science assessment. Hence, this article attempts to give an overview of the different assessment strategies that teachers can choose from to better evaluate science learning in their pupils.

Different strategies for the assessment of science learning

The assessment strategies that are discussed involve more of the continual assessment strategies to allow teachers to understand the progress of their pupils. These strategies have common features that differ from those of traditional strategies. First, they are less judgmental, and are more descriptive in the information that they provide to both teachers and learners on avenues for improvement. Second, they are not concerned solely with correct or incorrect answers, but emphasize more on how well pupils perform. These strategies provide a general picture of what pupils understand, what they are able to do, and how they apply the knowledge that they have learned.

There are a variety of strategies and opportunities for teachers to choose from in the measuring the progress of different aspects of the science learning of individual pupils, some of which are more appropriate than others, depending on the area of science that is being covered and the age range of the pupils (Hollins & Whitby, 1998). The assessment strategies that are available to assess the science learning of primary pupils include performance-based assessment in science projects and investigations, science journal writing, concept maps, portfolios, and questions and answers. Hughes and Wade (1996) suggest that it is important that a variety of methods should be used, because pupils may demonstrate their abilities differently with different approaches. For example, some pupils may perform better in "public" tasks such as oral discussion, and others may do better in "private" tasks, such as writing.

a) Performance-based assessment in science projects and investigations

The message that science is not only a body of knowledge but also a way of working seems to have reached teachers, but has not yet trickled down to pupils (Goldsworthy & Feasey, 1994). Although the processes of science are stressed, the continuous emphasis on subject knowledge in assessment has not allowed pupils to grasp the equal importance of science knowledge and science processes.

Science investigations and projects require pupils to explore science issues that they are interested in, or to apply science knowledge in designing things or finding ways to solve problems in everyday life. Diffily (2001) suggests that any science topic can become the focus for an investigation or a project. Any group of elementary pupils can learn to come to a consensus about a topic to study, conduct research, make day to day decisions about locating resources, organize what is being learned, and select a way of sharing with others what they have learned. Farmery (1999) recommends that investigations should be chosen carefully for primary school pupils. They should be adequately resourced, be easily adaptable, and be relevant to the curriculum so that they are assessable.

So and Cheng (2001) find that the multiple intelligences of Hong Kong primary pupils are developed through science projects. Active participation in science projects can help to sharpen the observation and thinking skills of pupils, cultivate their creativity, strengthen their exploration and analytical skills, facilitate their understanding of the relationship between science, technology, and society, and promote their desire to invent and explore.

Reinhartz and Beach (1997) suggest that it is often helpful to develop a set of criteria, or a grading rubric, for the evaluation of the responses and performance of pupils with performance-based assessment tools. The two performance-based assessment rubrics that are suggested in Demers' (2000) article are here merged to provide a clear picture of how the progress of pupils in observation skills, classification, and other areas of performance might be assessed (Table 1).

Level of performance	Process of inquiry	Evidence of inquiry	Depth of understanding	Communication	Presentation
High	Observations show evidence of careful study using multiple senses when appropriate. Descriptions contain intricate details. Classification systems clearly reflect careful observations made. System includes all samples provided	Questions are clearly identified and formulated in a manner that can be researched. Evidence and explanations have a clear and logical relationship. Methods of study generate valid data that addresses the question. Variables are controlled. Conclusions are based upon results and clearly explained.	Scientific information and ideas are accurate and thoughtfully explained. Patterns and trends are identified, discussed, and extended through interpolation and extrapolation.	Scientific information is communicated clearly and precisely and may include expressive dimensions. Presentation is effectively focused and organized.	Sentences are both complex and grammatically correct. Core words are spelled correctly. Punctuation is used appropriately. Script is neat and easy to read.
↓	Observations show evidence of careful study but are relegated to one sense. Descriptions are clear enough for samples to be accurately identified by another scientist. Classification systems are based upon the observations made.	Questions are clearly identified. Evidence and explanations have a logical relationship. Methods of study generate data that is related to the question. Variables are controlled. Conclusions are based on results.	Scientific information and ideas are accurate. Patterns and trends are identified.	Scientific information is communicated clearly. Presentation is focused and organized.	Sentences are grammatically correct. Most of the words, including the core words, are spelled correctly. Punctuation is used appropriately. Script is easy to read.
	Observations reflect the obvious characteristics of samples provided. Descriptions lack intricate detail. Classifications do not necessarily reflect observations made, and may not include all samples provided.	Questions are implied. Evidence and explanations have an implied relationship. Methods generate data related to the question. Variables are not controlled. Conclusions are related to the data.	Scientific information has occasional inaccuracies or is simplified. Patterns and trends are suggested or implied.	Scientific information has some clarity. Presentation has some organization and focus.	Sentences make sense but may contain grammatical errors. Text includes frequent spelling and punctuation errors. Script is legible.
Low	Observations lack clarity and detail, and are not clear enough to be interpreted by another scientist. Classification system is not based on observable characteristics and does not include all of the samples provided.	Questions are unclear or absent. Evidence and explanations have no relationship. Methods generate questionable data. Variables are not controlled. Conclusions are unclear or unrelated to the data.	Scientific information has major inaccuracies or is overly simplified. Patterns and trends are unclear or inaccurate.	Scientific information is unclear. Presentation lacks focus and organization.	The number of incomplete sentences and grammatical errors render the text difficult to interpret. Spelling and punctuation errors are prevalent. Script is illegible.

Table 1: Performance-based assessment rubric (Source: Demers, 2000, pp. 27-28)

Teaching for progression in experimental and investigative science is very difficult (Crossland, 1998). Crossland attempts to show how an aide-mémoire, laid out on one side of A4 paper, helps teachers to focus their short-term planning in terms of the curriculum and formative assessment. At the same time, he also shows the "pupil contribution" component, which provides very useful guidelines for the assessment of the progression of pupils in experimental and investigative science (Table 2). In addition, Farmery (1999) explains the development of a model for ensuring progression in experimental and investigative science. Table 3 shows an extract from this model that demonstrates the possible progression in "obtaining evidence" in experimental and investigative science.

Level	**Pupil Contribution**
Level 1 Observe using senses, talk, and draw	Help the teacher... Observe... Describe... Talk about...
Level 2 Make comparisons between observations and expectations	Respond to suggestions and, with help, make their own... Use the simple equipment provided ... Describe and compare ... Record (in simple tables)... Compare results with expectations...
Level 3 With some help, carry out a fair test	Respond to suggestions and make their own ... Simple predictions that can be tested... Measure using a range of equipment... With some help, carry out a fair test... Record in a variety of ways... Explain observations and any patterns arising out of their results ... Say what they found out...
Level 4 Recognize the need to carry out a fair test through description or action	Make predictions with a reason based on similar experience ... Recognize the need for a fair test by descriptions or actions ... Select equipment ... Make a series of observations/measurements adequate for the task... Present findings clearly in tables and bar charts... With help, plot graphs to find patterns and to relate conclusions to scientific knowledge and understanding...
Level 5 Carry out a scientific test in simple contexts involving only a few factors	Identify the crucial factors... Prediction based on scientific knowledge and understanding... Select and use apparatus with care... Measure and record with care... Ongoing interpretation of the results... Present data as line graphs... Draw conclusions that are consistent with the evidence... Begin to relate conclusions to scientific knowledge and understanding...

Table 2: Pupils' contribution in experimental and investigative science (Source: Crossland, 1998, p.19)

Obtaining Evidence			
Level 1	**Level 2**	**Level 3**	**Level 4**
Children use familiar equipment independently, e.g., weighing scales, ruler. They use unfamiliar equipment with support.	Children use a range of equipment independently, e.g., ruler, tape, weighing scales, thermometer.	Children use a wide range of equipment independently and accurately.	Children use a range of equipment, including instruments with fine divisions.
Children make observations of at least one event within each part of the investigation.	Children understand the need for detailed observations.	Children begin to realize the need to make a series of observations.	Children make a series of relevant and detailed observations
Children make measurements with some degree of accuracy.	Children identify and take relevant measurements.	Children recognize the need to make a series of measurements.	Children make a series of measurements with increasing precision.
Children occasionally repeat measurements to check if they have the same results as someone else.	Children recognize the need to repeat measurements to check accuracy.	Children repeat measurements to check accuracy.	Children repeat observations and measurements and begin to offer simple explanations for any differences recorded.

Table 3: Pupils' progression in "obtaining evidence" in experimental and investigative science (Source: Farmery, 1999, p.14)

b) Science journal writing

Pupils record procedures and results from investigations and observations, hypotheses, and inferences about science phenomena (Lowery, 2000). Free writing and drawing can also be used when the concept area involves possible long-term changes, and pupils should make regular observations (Hollins and Whitby, 1998). By creating journals, pupils are able to depict their way of seeing and understanding phenomena through their own lens of experience (Shepardson, 1997). The value of drawing and writing science lies in its potential to assist pupils to make observations, remember events, and communicate understandings (Shepardson & Britsch, 2000). Hollins and Whitby (1998) find that drawings and diagrams in response to a particular question are particularly revealing and informative when pupils add their own words to them, that is, annotations can help to clarify the ideas that a drawing represents.

Science journal writing with writing or drawings captures a dimension of conceptual understanding that is different from other types of assessment. Science journals can serve as diagnostic tools for informing practice, because they convey the understanding of pupils and so provide a window through which to view this understanding (Doris, 1991).

Shepardson and Britsch (1997) examine the ways in which science journals serve as a tool for teaching, learning, and assessment. They also discuss what science journals can say about what pupils are learning (Shepardson & Britsch, 2000). However, Shepardson and Britsch find that journals that were written by pupils on the topic of mixing and separating five different materials - clay, silt, sand, pebbles and gravel - give no indication that they have understood why sand and pebbles could be mixed and separated, and only show that it happened. Thus, journal writing might only indicate that pupils have learned the activity but not that they have learned the science. Therefore, Shepardson and Britsch remind teachers to employ multiple modes of assessment.

The ways that are suggested by Shepardson and Britsch (2000) to assess journals are simplified to help teachers to use journals as a meaningful tool for the assessment of the science learning of pupils.

- Determine whether pupils are representing the science activity or an understanding of the science.
- Look for differences between content understanding and science processes.
- Note which medium the primary pupil uses (i.e., drawing or writing).
- Look for details that indicate an understanding of the characteristics of objects or phenomena.
- Look at the ways in which the graphic context indicates the development level of pupils.
- Note the grammatical complexity of the writing.

The assessment logs in Table 4, which have been adopted and modified from Shepardson and Britsch (2000), can be used by teachers to monitor the performance of pupils in journal writing and drawing skills.

Assessment logs	Pupil performance
Representing science activity/understanding science	Activity/science understanding
Content understanding/science processes	Content/processes
Drawing/writing	Drawing/writing/drawing and writing
Graphic context	Labeling/immediate observation
Grammatical complexity of the writing	Description/analogy/explanation

Table 4: Assessment logs to monitor the performance of pupils in journal writing and drawing skills (Source:Adopted and modified from Shepardson & Britsch, 2000, p.32)

c) Concept maps

The use of concept maps in teaching and learning was initiated and developed by Novak and Gowin (1984). Concept maps measure or reflect more complex levels of thinking in the same way that science journals, science projects, science investigations, and other performance-based assessment methods do. In comparison with other assessment methods, however, concept maps are quicker, more direct, and considerably less verbal than essays or other types of written work. The visual nature of concept maps helps pupils to organize their conceptual framework (Willerman & MacHarg, 1991). White and Gunstone (1992) note that concept maps portray a great deal of information about the quality of learning and the effectiveness of the teaching. Stow (1997) states that concept mapping is a useful tool to help pupils to learn about the structure of knowledge and the process of knowledge production or meta-knowledge.

The use of concept mapping as an elicitation and assessment tool has been widely discussed (Atkinson & Bannister, 1998). Concept mapping has been shown to allow links to be made between concepts, and thus reveals scientifically correct propositions and misconceptions. The concept maps that are devised by pupils reflect their own ideas and understanding, and so cannot be marked wrong or right (Comber & Johnson, 1995), even if their ideas do not match with what is regarded as scientifically correct. Atkinson and Bannister (1998) have discovered that concept mapping can be a useful assessment tool, even with very young children.

By looking at the maps that were drawn by the pupils in Stow's article (1997), it is possible to see how the understanding of mapping and of the water cycle topic that is the subject of the maps has developed in the pupils. One pupil drew a fairly well connected map before the investigations (Figure 2) that seems to show a mixed understanding of the concepts involved. After the investigations, the same pupil's map is significantly more sophisticated (Figure 3), and shows a far greater range of connections and a greater understanding of the grammar that is needed to complete the connections. It demonstrates a clearer understanding of the concepts that are involved; for example, evaporation is linked to condensation, and also to the sun. The motivational benefits of the comparison of the two maps and the pupil's self-evaluation of their progress are clear. The opportunity that concept mapping provides for pupils to examine the progress of their own learning is instrumental in the encouragement of meaningful learning. The mapping and subsequent evaluation provides a framework of reference within which pupils can analyze their own thinking, which enables them to identify their strengths and weakness and set themselves future learning targets.

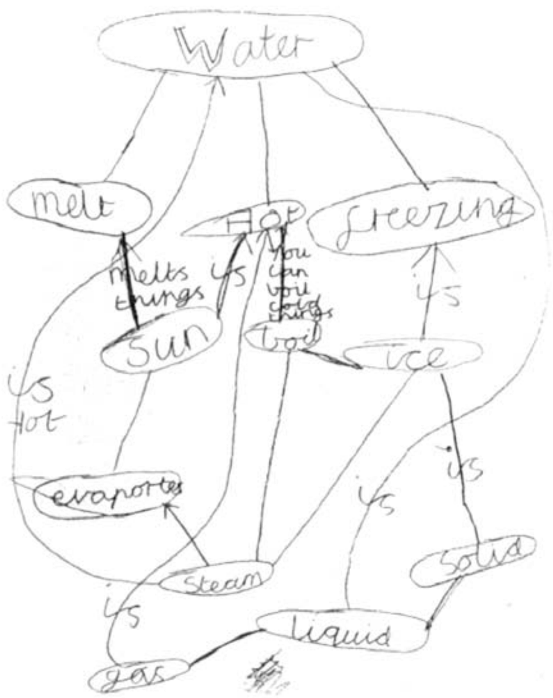

Figure 2: A pupil's concept map before carrying out the activity (Stow, 1997, p.13)

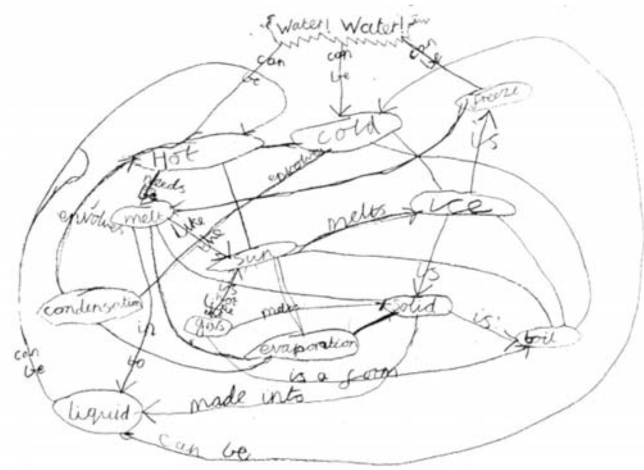

Figure 3: A pupil's concept map after carrying out the activity (Stow, 1997, p.13)

Concept maps serve both formative and summative purposes in the assessment of student science learning. Over the past twenty-five years, concept maps have been adopted by thousands of teachers in elementary and secondary schools (Edmondson, 1999). The following are comments from science educators on the advantages of using concept maps as assessment tools.

- The concept map can be read quickly and can show a large body of information in a concise and clear manner (Comber & Johnson, 1995).
- The advantage of concept maps is that they are formative, and can be completed quickly (Hollins & Whitby 1998).
- Concept maps may be used in classroom activities to provide students with immediate feedback about the depth of their understanding, or to assess learning from specific instructional units that might not otherwise be reflected by paper and pencil tests (Markow & Lonning, 1998).
- Concept mapping is unique in comparison to traditional achievement tests, the limitations of which render them inadequate for tapping certain characteristics of knowledge structures (Hoz, Tomer, & Tamir, 1990).
- Trowbridge and Wandersee (1996) use concept maps to analyze differences in the comprehension of pupils and find concept mapping to be a highly sensitive tool for measuring changes in knowledge structure, and particularly for depicting changes in pupils' selection of super ordinate concepts.
- Wallace and Mintzes (1990) use concept maps to document conceptual change in biological concepts, and the concept maps of the pupils in their study reveal significant and substantial changes in the complexity and prepositional structure of their knowledge base.
- Concept maps are particularly helpful in representing qualitative aspects of learning. They may also be used by teachers to evaluate learning. They are meta-cognitive tools

> that can help both teachers and pupils to better understand the content and process of effective, meaningful learning (Edmondson, 1999). Concept maps are tools for representing the interrelationship between concepts in an integrated, hierarchical manner.

Novak and Gowin (1984) suggest that teachers could construct a "criterion map" against which the maps of pupils could be compared, and the degree of similarity between the maps could then be given a percentage score. However, White and Gunstone (1992) argued that though scoring is not helpful for formative assessment, scoring becomes more sensible when concept maps are used in summative assessment. There are various other schemes for scoring concept maps, but most of them are variations of the scheme that is outlined by Novak and Gowin. Markham, Mintzes, and Jones (1994) modified Novak and Gowin's scheme to include three more observed aspects of concept maps for scoring: the number of concepts, which is evidence of the extent of domain knowledge; concept relationships, which provide additional evidence of the extent of domain knowledge; and branching, which provides evidence of progressive differentiation. Table 5 shows a summary of the schemes that are suggested by Novak and Gowin (1984) and Markham, Mintzes and Jones (1994).

There are other suggestions for the scoring of concept maps. Trowbridge and Wandersee (1996) suggest a concept map "performance index," which they describe as a compound measure that includes the pupil's concept map scores, the difficulty level of each map produced, and the total number of maps submitted. Rice, Ryan, and Samson (1998) developed a method of scoring concept maps that is based on the correctness of the propositions that are outlined in a table of specifications of instructional and curriculum goals. They find high correlations between concept map scores and scores in multiple-choice tests that are aimed at assessing the same instructional objectives. Edmondson (1999) suggests that the scores for particular attributes of concept maps could be used as a basis for a comparison of the extent to which different dimensions of understanding have been achieved. The purpose of such assessment is for teachers to make adequate provision for pupils' learning to further develop their understanding.

Novak and Gowin (1984)		**Markham, Mintzes & Jones (1994)**	
Criteria for evaluating concept maps	Scoring	Criteria for evaluating concept maps	Scoring
Validity of relationship	1 point for valid relationship	Concept relationships that provide additional evidence of the extent of domain knowledge	1 point for each valid relationship
Levels of hierarchy	4 points for hierarchy	Hierarchies that provide evidence of knowledge	5 points for each level of hierarchy
Validity of the propositions and cross-links	10 points for each cross-link	Cross-links that represent evidence of knowledge integration	Each cross-link receives 10 points
Use of examples	1 point for each example	Examples that indicate the specificity of domain knowledge	Each example receives 1 poin
		Number of concepts as evidence of the extent of domain knowledge	1 point for each concept
		Branching, which provides evidence of progressive differentiation	1 point for each branching, 3 points for each successive branching

Table 5: A summary of the schemes for scoring concept maps

Although there are many suggestions for the scoring of concept maps, there are also criticisms of these scoring systems. Regis, Albertazzi, and Roletto (1996) therefore suggest a shift in emphasis and focus toward the assessment of changes in the content and organization of concept maps over time.

d) Portfolio

Spandel (1997) asserts that any collection of student work, which includes tests, homework, and laboratory reports, can be included in a portfolio as representative samples of student understanding. Portfolios provide examples of individual student work, and can indicate progress, improvement, accomplishment, or special challenges (Lowery, 2000). Portfolios should be a collection of many meaningful types of materials that provide tangible proof of the progress of a pupil (Reinhartz & Beach, 1997). As part of the portfolio exercise, Buck (2000) has pupils pick out their best work from a unit and describe what the pieces of work reveal about what they have learned. Vitale and Romance (2000) focus on the value of portfolios as measures of understanding in natural science, and further suggest that portfolios might be defined as collections of student work samples that are assumed to reflect the meaningful understanding of the underlying science concepts (Vitale & Romance, 2000). They highlight that portfolio activities and tasks are open-ended, and constructively require pupils to use and apply knowledge in ways that demonstrate their understanding of science concepts.

Portfolios are one of the assessment measures that were recommended in the recent curriculum reform in Hong Kong: "The portfolio is used to contain students?evidence of learning. During the processes, pupils make their own judgment and select the artifacts (observation sheets, questionnaire and interview results, art produced, etc.) that best meet the criteria for excellence and personal improvement?(Curriculum Development Council, 2000, p.16). As a form of authentic assessment, portfolios are considered by their advocates to offer teachers a more valid means of evaluating student understanding than traditional forms of testing (Jorgenson, 1996).

Within science classrooms, a wide range of products can be included as work examples in student portfolios. The emphasis should be on products that reflect the meaningful understanding, integration, and application of science concepts and principles (Raizen, Baron, Champagne, Haertel, Mullis & Oakes, 1990). These include reports of empirical research, analyses of societal issues from a sound scientific view, papers that demonstrate an in-depth understanding of fundamental science principles, the documentation of presentations that are designed to foster the understanding of science concepts for others, journals that address a pupil's reflective observations over an instructional time span, and analytic or integrative visual representations of science knowledge itself in the form of concept maps.

Vitale and Romance (2000) suggest the development of guidelines for the evaluation of portfolio assessment products. The evaluation of the portfolio by the teacher should be a clinical judgment with two considerations. The first is the degree to which the relevant conceptual knowledge is represented accurately in the portfolio product, and the second is the degree to which the portfolio product meets the specified performance outcomes, which include the degree to which the relevant concepts are used on an explanatory or interpretative basis by pupils. Thus, there is no need to develop numbered scoring systems or rubrics, because they are not specific enough to provide evidence of meaningful student learning.

e) Questions and Answers

Open-ended questions mimic good classroom strategies and encourage thinking (Lowery, 2000), both of which are helpful to teachers in understanding how pupils go about finding an answer, solving a problem, or drawing a conclusion. Hughes and Wades (1996) also suggest that both open-ended and closed questions might be asked to gain information about pupils' investigational abilities. Some examples of open questions are:

- What questions are likely to be asked about the cars and the slopes?
- How did you ensure that you carried out a fair test?
- How do your observations compare with your prediction?

Hughes and Wades (1996) acknowledge the flexible nature of one to one or group questioning. These techniques enable supplementary questions to be asked to clarify what was really meant by a child's vaguely worded response or to verify whether omitted details from a written account were due to forgetfulness, laziness, or a lack of understanding and ability.

However, Black and Wiliam (1998) opine that the dialogue between teacher and pupils that arises when the teacher asks questions is unproductive in the assessment of learning. There may be a lowering of the level of question to facts that require very little thinking time, and the dialogue only ever involves a few pupils in the class. To enable thoughtful, reflective, and focused dialogue between teacher and pupils to evoke and explore understanding, Black and Wiliam (1998) suggest that teachers should:

- give pupils time to respond, ask them to discuss their thinking in pairs and in small groups, and ask a representative to speak on behalf of a small group;
- give pupils a choice between different possible answers and ask them to vote on the options; and
- ask all of the pupils to write down an answer and then read out a selected few.

In addition to questions from teachers, Watts, Barber, and Alsop (1997) assert that the questions of pupils can be very revealing about the way that they think, their worries and concerns, what they want to know and when they want to know it. Gibson (1998) uses a similar technique in his study, but with an emphasis on the answers of pupils to their own questions. The process of asking questions is emphasized, and the construction of meaning should be continuous. Asking pupils to generate questions on a regular basis also shows their development, as the questions really start to probe the big issues, or narrow the topic down to very specific queries. Gibson shows a sample of the range of pupils' answers in her article. The small selection of pupils' explanations clearly shows their thinking, and is possibly even more revealing than their questions. Gibson states that the answers that pupils give to their own questions can be a valuable learning and assessment tool. Although some of the pupils in his study showed a shallower understanding than others, the answers of all of the pupils give an insight into how they are developing. This "Any answer" session that is suggested by Gibson can follow on from sessions that are designed to generate questions, either before or after practical investigations, and will reveal the thinking of the pupils and their ability to make a hypothesis.

The important qualities of meaningful assessment

Although the suggested assessment strategies provide information about science learning that paper and pencil assessments are not able to provide, these formative assessment strategies may take time, and may look very different from the ones with which pupils and parents are familiar (Lowery, 2000). Moreover, Black (2000) finds that the effectiveness of formative feedback depends upon several detailed features of its quality, and not on its mere existence or absence. As advocated by science educators and researchers, interactive assessment, coherent assessment systems, self-assessment, peer assessment, and feedback are the most important qualities of the new types of assessment. Figure 4 shows the relationship of these important qualities in the meaningful assessment of learning.

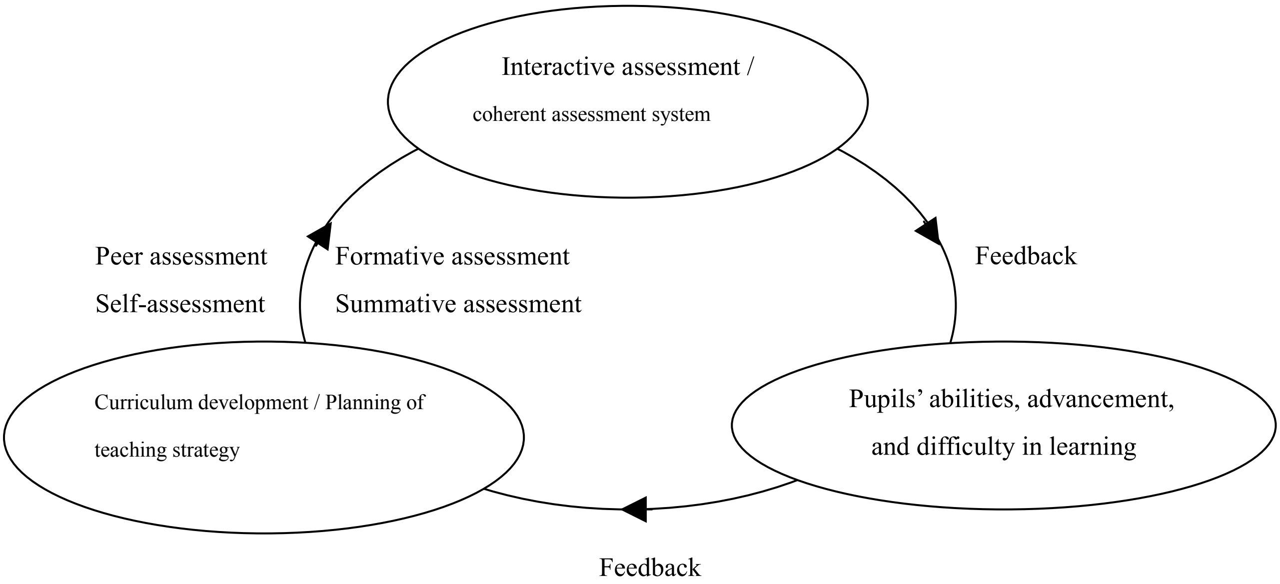

Figure 4: The relationship of the important qualities of the meaningful assessment of learning

Interactive assessment

There are two types of formative assessment in Cowie and Bell's (1999) model: planned formative assessment and interactive assessment. The process of planned formative assessment is characterized by the teacher eliciting, interpreting, and acting on assessment information. Interactive formative assessment takes place during student-teacher interaction, which involves teachers noticing, recognizing, and responding to student thinking during this interaction. The details of interactive assessment are not planned (Bell, 2000) and cannot be anticipated, but teachers need to be prepared for interactive formative assessment, and to use teaching approaches that allow it to occur naturally.

Coherent assessment system

The information from any one assessment is one piece of the puzzle that makes up instruction (Lowery, 2000). Reinhartz and Beach (1997) further elaborate that assessment has many parts and is a gigantic puzzle that has many separate pieces. The challenge is to fit all these pieces together into a coherent assessment system. A coherent assessment system means the adoption of various methods to measure progress towards the achievement of the goals and objectives of science instruction. Atkinson and Bannister (1998) suggest that the solution to a coherent assessment system should not rely on a single form of assessment tool, as pupils can demonstrate varying levels of understanding when they respond in different modes.

Self-assessment and peer assessment

It is clear that effective formative feedback can be all pervasive (Black, 2000). Feedback takes place through written and oral questions, through the quality of classroom dialogue, and through the formulation of classroom tasks so that pupils are active enough to produce feedback evidence. However, it is also important to involve the pupils through the medium of peer assessment and self-assessment.

Peer assessment involves the cross-referencing of the evidence of the contribution of individuals to the completion of group tasks, and self-assessment involves the assessment of one's own progress through the work process. Reinhartz and Beach (1997) state that peer assessment is important because cooperative and group work is an integral part of science learning. Peer assessment means making judgments based on individual responsibilities that are performed for the benefit of the group. Moreover, in peer assessment, pupils are aware that they are communicating to an audience, and that writing and reporting is therefore undertaken for a purpose (Lindsay & Clarke, 2001).

Lindsay and Clarke (2001) recognize and appreciate the advantages of self-assessment in the enhancement of learning. With self-assessment, pupils can use the skills of self-marking in a variety of contexts, and thus develop greater perseverance in their learning. Lindsay and Clarke further elaborate that self-marking can clarify ideas for pupils, and can help them to refine and question their own concepts. Self-assessment encourages pupils to be constantly involved in the scientific process and their role within it, and helps them to become more scientific in their enquiries. Furthermore, through self-assessment, pupils raise questions that constantly reinforce their understanding of the skills and knowledge that they are acquiring. This encourages pupils to open up new avenues of investigation.

Black and Wiliam (1998) argue that the main problem of self-assessment is that pupils can assess themselves only when they have a sufficiently clear picture of the targets that their learning is meant to reach. However, many pupils do not possess such a picture, and thus pupils should be trained in self-assessment to acquire an overview of their learning and thereby grasp what they need to do to achieve the target. Self-assessment often appears to be problematic, but it can be effective given clear guidelines. It may be combined with peer assessment and teacher assessment, or used on its own. Decisions then have to be made on whether, and in what proportion, self- and peer assessment becomes a part of the final grade. Whatever the form, these modes of assessment ensure a rich learning experience (Biggs, 1995).

Feedback

Black and Wiliam (1998) suggest that tests that are given in class, and other exercises that are assigned for homework, are important means of promoting feedback. Leakey and Goldsworthy (2001) state that feedback is what teachers should aim to give pupils as part of the assessment process. However, research studies have shown that if pupils are given only marks or grades, then they do not benefit from assessment. It is thus concluded that feedback on tests and other assessment tasks should give pupils guidance on how to improve, and individual pupils should be given the assistance and opportunity to work toward improvement. Leakey and Goldsworthy (2001) highlight that the task of making the feedback meaningful and valuable can be complex and demanding, and that constructive feedback should involve the sharing of learning objectives, the advancement of the pupil thinking, and the fomentation of an ethos of trust between the teacher and the learner.

Conclusion

The use of summative paper and pencil tests in the assessment of science learning is less time consuming and causes little disturbance to teachers or parents, and the results that are reflected in the report cards of pupils are familiar to everyone. However, by using challenging and exciting alternative formative assessment methods, teachers will be able to perceive the development of meaningful modes of inquiry and science understanding by pupils (Lowery, 2000). Moreover, formative assessment has the potential to drive changes in teaching that can improve the conceptual understanding of pupils dramatically, and teachers necessarily must alter the way in which they evaluate this learning and formative assessment to effect this change (Dougherty, 1997). It is thus recommended that both summative assessment and formative assessment should be used in equal measures to periodically evaluate the cumulative understanding and continuing progress of pupils in their learning.

There is a clear message that teachers have a variety of alternative types of assessment from which to choose. Teachers may opt to combine different methods in the evaluation of different aspects of an individual pupil's progress. The different pathways to better assessment may be new to some teachers, it is hoped that the information is useful to help primary school teachers to meaningfully assess their pupils.

References

Atkinson, H. & Bannister, S. (1998). Concept maps and annotated drawings, *Primary Science Review*, 51, 3-5.

Bell, B. and Gilbert, J. (1996). *Teacher Development: A Model from Science Education*. London: Falmer Press.

Bell, B. (2000) Formative assessment and science education: a model and theorizing. In R. Millar, J. Leach & J. Osborne (Eds) *Improving Science Education: the Contribution of Research*, 48-61. Philadelphia: Open University Press.

Biggs, J. (1995). Assessment of learning. In J. Biggs & D. Watkins (Eds) *Classroom Learning: Educational Psychology for the Asian Teacher*. 167-194. Singapore: Prentice.

Black, P. & Wiliam, D. (1998). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan*, *(80)2*, 139-148.

Black, P. (2000). Policy, practice and research: the case of testing and assessment. In R. Millar, J. Leach, & J. Osborne (Ed.) *Improving Science Education: the Contribution of Research*, 327-346. Philadelphia: Open University Press.

Buck, G. A. (2000). Teaching science to English-as-second-language learners, *Science and Children, (38)3*, 38-41.

Comber, M. & Johnson, P. (1995). Pushes and pulls: the potential of concept mapping for assessment. *Primary Science Review*, 36, 10-12.

Cowie, B. & Bell, B. (1996). Validity and formative assessment in science education, paper presented to the *Symposium on Validity in Educational Assessment, Dunedin, New Zealand, 28-30*.

Cowie, B. & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education, 6(1),* 101-106.

Crossland, J. (1998). Teaching for progression in experimental and investigative science, *Primary Science Review*, 53, 18-19.

Curriculum Development Council (2000). *Learning to Learn: General Studies for Primary Schools (Consultation Document)*. HKSAR Printing Department.

Daws, N. & Singh, B. (1996). Formative assessment into what extent is its potential to enhance pupils science being realized? *School Science Review, 77(281)*, 93-101.

Demers, C. (2000). Beyond paper and pencil assessment, *Science and Children (38)3*, 24-29.

Diffily, D. (2001). Project Reptile! *Science and Children, (38)7*, 30-35.

Doris, E. (1991). *Doing What Scientists Do: Children Learn to Investigate Their World*. Portsmouth, NH: Heinemann.

Dougherty, M. J. (1997). Formative assessment: using an instructional model to improve conceptual learning. *The Science Teacher, 64(6)*, 29-33.

Driver, R. (1989). Students' conceptions and the learning of science, *International Journal of Science Education, 11(5)*, 481-90.

Edmondson, K. M. (1999). Assessing science understanding through concept maps. In J. D., Mintzes, J. H. Wandersee, & J. D. Novak (Eds) *Assessing Science Understanding: a Human Constructivist View*, 22-33. Toronto: Academic Press.

Farmery, C. (1999). Progression in experimental and investigative science, *Primary Science Review, 58,* 12-14.

Gibson, J. (1998). Any questions any answer? *Primary Science Review, 51,* 20-21.

Goldsworthy, A. & Feasey, R. (1994). *Making sense of primary science investigation.* Herts: Association of Science Education.

Glynn, S. G., Yeany, R. H., & Britton, B. K. (1991). A constructivist view of learning science. In S, M. Glynn, R. H. Yeany., & B. K. Britton. (Eds) *The Psychology of Learning Science*, 3-19, New Jersey: Lawrence Erlbaum Associates, Inc.

Hayes, P. (1998). Assessment in the classroom. In R. Sherrington (Ed.) ASE *Guide to Primary Science Education, 125-129*. UK: Association for Science Education.

Hollins, W. & Whitby, V. (1998). *Progression in Primary Science*. Great Britain: David Fulton Publishers.

Hoz, R., Tomer, Y., and Tamir, P. (1990). The relations between disciplinary and pedagogical knowledge and the length of teaching experience of biology and geography teachers, *Journal of Research in Science Teaching*, 27(10), 973-985.

Hughes, C., & Wade, W. (1996). *Inspirations for Investigations in Science*. Warwickshire: Scholastic.

Jorgenson, M. (1996). *Rethinking Portfolios Assessment: Document the Intellectual Work of Learners in Science and Mathematics.* ERIC Document Reproduction Services [ED400169].

Leakey, A. & Goldsworthy, A. (2001). Fantastic feedback. *Primary Science Review, 68*, 22-23.

Lee, O. & Fradd, S. H. (1998) Science for all, including students from non-English language backgrounds, *Educational Researcher*, 27(4), 12-21.

Lindsay, C. & Clarke, S. (2001). Enhancing primary science through self- and paired assessment. *Primary Science Review, 68*, 15-18.

Lowery, L. F. (2000). *NSTA Pathways to the Science Standards*, Arlington: National Science Teacher Association.

Markham, K. M., Mintzes, J. J., & Jones, M. G. (1994). The concept map as a research and evaluation tool: further evidence of validity. *Journal of Research in Science Teaching, 31(1)*, 91-101.

Markow, P. G. & Lonning, R. A. (1998). Usefulness of concept maps in college chemistry laboratories: students?perceptions and effects on achievement. *Journal of Research in Science Teaching, 35(9),* 1015-1029.

Novak, J. & Gowin, D. (1984). *Learning to Learn*. Cambridge: Cambridge University Press.

Novak, J. D., Mintzes, J. J., & Wandersee, J. H. (1999) Learning, teaching and assessment: a human constructivist perspective. In J. D., Mintzes, J. H. Wandersee, & J. D. Novak (Eds) *Assessing Science Understanding: a Human Constructivist View,* 1-13, Toronto: Academic Press.

Osborne, R. J. & Wittrock, M. (1985). The generative learning model and its implications for learning in science, *Studies in Science Education,* 12, 59-87.

Raizen, S. A., Baron, J. B, Champagne, A. B., Haertel, E., Mullis, I. V. S., & Oakes, J. (1990). *Assessment in Science Education: the Middle Years.* Washington, DC: National Center for Improving Science Education.

Regis, A., Albertazzi, P.G., & Roletto, E. (1996). Concept maps in chemistry education. *Journal of Chemistry Education, 73(11)*, 1084-1088.

Reinhartz, J. & Beach, D. M. (1997). *Teaching and Learning in the Elementary School: Focus on Curriculum*. New Jersey: Prentice-Hall.

Rice, D. C., Ryan, J. M., & Samson, S. M. (1998). Using concept maps to assess student learning in the science classroom: must different methods compete? *Journal of Research in Science Teaching, 35(10)*, 1103-1127.

Shepardson, D. P. (1997). Butterflies and beetles: first graders' ways of seeing and talking about insect life cycles. *Journal of Research in Science Teaching, 34(9)*, 876-889.

Shepardson, D. P. & Britsch, S. J. (1997). Children's science journals: tools for teaching, learning, and assessing. *Science and Children, 34(5),* 13-17.

Shepardson, D. P. & Britsch, S. J. (2000). Analyzing children's science journals, *Science and Children*, (38)3, 29-33.

So, W. M. W. & Cheng, M. H. M. (2001). To facilitate the development of multiple intelligences among primary students through science projects. *Asia-Pacific Forum on Science Learning and Teaching, 2(1), Article 4*. Available at: http://www.ied.edu.hk/apfslt/v2_issue1/sow/ .

Spandel, V. (1997). Reflections on portfolios, in G. D. Phye (Ed.), *Handbook of Academic Learning*, 573-591, San Diego, CA: Academic Press.

Stow, W. (1997). Concept mapping, a tool for self-assessment. *Primary Science Review, 49*, 12-15.

Trowbridge, J. E. & Wandersee, J. H. (1996). How do graphics presented during college biology affect students learning? *Journal of College Science Teaching, 26(1)*, 54-57.

Vitale, M. R., & Romance, N. R. (2000). Portfolios in science assessment: a knowledge-based model for classroom practice. In J. D., Mintzes, J. H. Wandersee, & J. D. Novak (Eds) *Assessing Science Understanding: a Human Constructivist View,* 167-196. Toronto: Academic Press.

Wallace, J. D. & Mintzes, J. J. (1990). The concept map as a research tool: Exploring conceptual change in biology. *Journal of Research in Science Teaching, 27(10),* 1033-1052.

Watts, M., Barber, B., & Alsop, S. (1997). Children's questions in the classroom, *Primary Science Review, 49, 6-8*.

White, R. and Gunstone, R. (1992). *Probing Understanding*. London: Falmer Press.

Willerman, M. & MacHarg, R. (1991). The concept map as an advance organizer. *Journal of Research in Science Teaching, 28(8),* 705-711.