Title: Does formal assessment in primary science support the TSLN vision?
Author(s): Boo Hong Kwen
Source: ERAS Conference, Singapore, 19-21 November 2003
Organised by: Educational Research Association of Singapore (ERAS)

This document may be used for private study or research purpose only. This document or any part of it may not be duplicated and/or distributed without permission of the copyright owner.

The Singapore Copyright Act applies to the use of this document.
DOES FORMAL ASSESSMENT IN PRIMARY SCIENCE SUPPORT THE TSLN VISION?

Boo Hong Kwen
National Institute of Education
Nanyang Technological University
Singapore

Abstract

The attainment of educational goals such as the inculcation of creativity, problem-solving and other higher order thinking needs to be supported in the teaching of specific content, appropriate methods of teaching as well as relevant modes of assessment. As a broader range of teaching methods has been developed in schools in the period from 1987 through to 1990s and the current millennium, with special impetus provided by the “Thinking Schools Learning Nation” drive, the question being raised in this paper is has assessment moved in tandem? This paper is focused on formal assessment at the primary science level and highlights some areas where assessment items may in fact be counter-productive to the TSLN vision.

Keywords: Primary science, assessment, higher order thinking, problem solving, TSLN

Background And Introduction

The concern to promote innovation and creativity, depth and breadth in thinking could be traced to 1987 when the teaching of Cort 1 (de Bono’s module on the teaching of Breadth in Thinking) and Cort 4 (de Bono, 1986) has been promoted in schools in Singapore. The move to teach creativity and higher order thinking skills explicitly to students continue to evolve and develop, and culminate in the Thinking Programme (Chua & Leong, 1998). This concern is consolidated under the “Thinking Schools, Learning Nation (TSLN)” drive to generate creativity and thinking across schools and in the nation (Goh, 1997) as well as the “Innovation and Enterprise” or “I&E” goal (Shanmugaratnam, 2003).

This author is concerned with the question, “Has formal school assessment kept in step with the change in teaching methods which emphasized the explicit teaching of creative and higher order thinking skills?” The focus of this paper is on formal science assessment in primary schools and is limited to a convenient sample of 100 sets of primary school semester assessment papers from 25 schools. Each school provided four sets of examination papers (P3-P6 papers). Each set of papers comprised typically 30 multiple choice type questions (MCQs) and 16 open-ended supply type or free response questions. Based on the scrutiny of these papers, the following four categories of questions which are regarded as unsupportive of, or counterproductive to, the TSLN vision are highlighted and discussed. It should be emphasized here that while there were questions which were flawed in these four particular ways, there were also well-crafted questions which effectively tested pupils’ higher order thinking skills and acquisition of science concepts and which were also fully aligned with the TSLN objectives.
Four Categories Of Questions That Are Unsupportive Of Creative Or Higher Order Thinking Skills

Category 1:
Questions that involve perceptual mismatch between question setter and pupils.

In some of the test items, there is considerable potential for perceptual mismatch between the question setter and the pupils. What may appear to be a well-bounded and precise question on the part of the question setter can often be interpreted quite differently by the pupils. This problem of mismatch in perception of the question posed is more pronounced where pupils who are capable of higher order or creative thinking are concerned. This is because these pupils tend to see issues, alternatives and ambiguity in test items that the setter did not intend or was unaware of. This is particularly severe in the case of multiple-choice questions (MCQs) where there is supposed to be one and only one correct answer out of four given options for each question and where pupils’ answers are marked by computer. In some MCQs (such as question 1.1 below), pupils who are thinking laterally or deeply, find situations in which all or none of the options can be correct, while pupils who are not as capable in their thinking or who know the material in the conventional way could simply select the most appropriate (often most obvious) answer and get marked correct.

Example Question 1.1

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Which one of the following animals should not be in the same group as the others?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>duck</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>goat</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>lion</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>mole</td>
</tr>
</tbody>
</table>

Comment

In this question, the setter’s answer key is option 1 “duck”. To the setter, who has taught pupils the classification of vertebrates into groups such as mammals, birds, fish, reptiles, amphibians and so forth, the given item is a very easy question, almost a give-away, since option 1 “duck” is clearly the odd-one out, the only bird among the three mammals given. However, because the basis of classification or grouping is not made explicit in the question, a pupil who is thinking laterally could opt for any of the other alternatives as the answer:

- Option 2 could be the answer key since “goat” is the only animal with horns.
- Option 3 could be the answer key since “lion” is the only carnivore.
- Option 4 could be the answer key since “mole” is the only nocturnal animal or the only animal that burrows and lives underground.

One approach to alleviate the problem of perceptual mismatch between question setter and pupils is to frame more tightly bounded questions to define more precisely the question scenario as originally perceived by the question setter. However, the additional qualification(s) may mean that the correct answer would be
made more obvious and hence easily selected by pupils, as could be seen from the improved version of Question 1.1 shown below.

*A suggested improved version of Question 1.1*

<table>
<thead>
<tr>
<th>Based on type of body covering, which one of the following animals should <strong>not</strong> be in the same group as the others?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 duck</td>
</tr>
<tr>
<td>2 goat</td>
</tr>
<tr>
<td>3 lion</td>
</tr>
<tr>
<td>4 mole</td>
</tr>
</tbody>
</table>

In this revised version, by specifying the basis of classification, the answer key has been made rather obvious, since it is common knowledge that duck has feathers as a body covering while the other three animals have hair.

**Example Question 1.2**

<table>
<thead>
<tr>
<th>Which one of the following animals is <strong>unlike</strong> the others in terms of how it moves?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 dog</td>
</tr>
<tr>
<td>2 frog</td>
</tr>
<tr>
<td>3 rabbit</td>
</tr>
<tr>
<td>4 kangaroo</td>
</tr>
</tbody>
</table>

**Comment**

In this question, unlike the previous one, the setter has specified the basis of classification in terms of movement of the animals. The answer key is option 1 “dog”. To the setter, who has taught pupils the topic of “animals and their movement”, the given item is a very easy one, almost a give-away, since the dog does not hop while the other animals do. However, pupils, especially the ones trained in creative and higher order thinking, probably could see alternatives not anticipated by the setter, and could choose any of the other options as the answer.

- Option 2 could be the answer key since the “frog” is the only animal that could swim underwater, being the only animal among the four given that capable of breathing through its skin and as well as its lungs. The other three animals could swim, if forced to, but with their heads above the water (by doing strokes known as “dog paddle” since they can only breathe through their lungs.
- Option 3 could be the answer key since the “rabbit” is the only animal that can burrow or move underground.
- Option 4 could be the answer key since the “kangaroo” is the only animal that uses only two legs when moving at optimum speed.

One way of addressing this kind of perceptual mismatch between the question setter and the examination takers is to use two-tier questions, an example of which is given below.
An example of a two-tier (improved) version of Question 1.2

Which one of the following animals is unlike the others in terms of how it moves?

1. dog
2. frog
3. rabbit
4. kangaroo

(a) Explain your answer.

Category 2:

Questions that are incorrectly focussed on low-level recall or extraction of information

Included in this category are questions in which the design of an investigation is given and the pupils are required only to use low level thinking such as extraction of information or recalling what they have learned in class to answer the questions. These kinds of questions do not encourage creative or higher order thinking, and are counterproductive to the TSLN impetus.

Example Question 2.1

Jo put four tomato seeds (A, B, C and D) under the conditions given below and observed their growth.

<table>
<thead>
<tr>
<th>Seed</th>
<th>Water</th>
<th>Air</th>
<th>Light</th>
<th>Temp/°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>30</td>
</tr>
<tr>
<td>B</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>30</td>
</tr>
<tr>
<td>C</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>29</td>
</tr>
</tbody>
</table>

(a) Which seed is most likely to germinate?
(b) What can you say about the conditions under which the seed will germinate?

Comment

That the required answer to part (a) “seed A” is obvious from the table since students have been taught in class that seeds need water, air and warmth in order to germinate. All the students need to do is to decide between seed A and seed C both of which were supplied with water and air, and with seed C being kept at the lower temperature of 2ºC compared to seed A’s 30ºC. The answer to (b) is based on recall of teacher’s drilling that “seeds need water, air and warmth in order to germinate”.

This kind of question not only does not encourage creative or deep thinking but also provides a negative example of the way to conduct an experiment that involves living things. In experiments dealing with organisms (whether plant or animal) a sample size of one (1) is unacceptable. Unlike physical science experiments where non-living things are involved, and where relationships between variables can often be described unambiguously (“proportional to”, “inversely proportional” or
defined in terms of an equation) living things such as seeds incorporate many uncontrollable variables and their behavior cannot be reliably predicted. The exploration of possible relationships between causes and supposed effects is not as straightforward as physical science experiments. Correlational reasoning is the process used to assess the strength of relationships between variables, and the sample size for each condition investigated should be as large as is feasible. It would have been better to re-frame or re-phrase the question in such a way that it allows students to exercise their higher order thinking skills and creativity by designing their own investigations, instead of being provided with the investigation. An example of a re-framed question is given below.

A suggested improved version of Question 2.1

Sally said, “Seeds need water, air and warmth in order to germinate.”
Outline an experiment to find out if any one of the three factors mentioned by Sally is correct. Assume you are given a bag of red beans and any other materials and apparatus you might need.

Such a re-framing or re-crafting of the question would, not only cater to the creative and problem-solving, higher order thinking skills of pupils better, but would also test the pupils’ understanding of the basic principles of good experimental designs, such as the correct identification of independent and dependent variables, control of appropriate variables and extraneous factors, use of appropriate sample size and the incorporation of experimental controls.

Example Question 2.2

Peter did an experiment to measure the force needed to start a wooden block moving over four different surfaces and some of the results are given in the table below.

<table>
<thead>
<tr>
<th>Type of surface</th>
<th>Sandpaper</th>
<th>Carpet</th>
<th>wood</th>
<th>glass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Force required/gf</td>
<td>60</td>
<td>90</td>
<td>?</td>
<td>35</td>
</tr>
</tbody>
</table>

What would be the force needed to move the wooden block over the wooden surface? (1) 30gf (2) 45gf (3) 65gf (4) 80gf
Comment

Unlike the previous example, this question 2.2 involves non-living things. While non-living things are more predictable, in any physical science experiment, there are various uncertainties or errors associated with each trial of the experiment. These uncertainties can be minimized by carrying out repeated trials for each variable being tested. Question 2.2 as it is does not exemplify good experimental procedure because there is no mention that Peter carried out repeated trials for each type of surface he tested.

Besides illustrating poor experimental design, this question is flawed also in terms of the setter’s assumption that the test takers would have in their minds exactly the same kind of carpet, sandpaper, wood and glass that the question setter himself/herself has. The fact is, there are so many different kinds of carpet (e.g. the roughness of which would depend on whether it is made of synthetic or natural material), different degrees of roughness among different grades of sandpaper, different kinds of wood (ranging from very rough to very smooth) and even different kinds of glass (with varying roughness). It is unreasonable to expect the pupils to have in their minds exactly the same kinds of these materials as the setter!

As in the previous example, this question should be re-crafted to test pupils’ creativity and higher order thinking/problem solving skills, so that it becomes more aligned with the TSLN vision. An example of a re-crafted version is given below.

A suggested improved version of Question 2.2

Outline an experiment to show that different amounts of force are needed to move a wooden block over different types of surfaces. Assume that you are provided with the following types of surfaces - wood, glass, carpet and sandpaper, and any other materials and apparatus that you might need.

Category 3: Questions which test trivial or esoteric facts

Questions that focus on trivial or esoteric facts rather than science concepts or processes can impede the fostering of creative and higher order thinking skills in pupils. Question 3.1 which simply requires pupils to recall the characteristics of cockroach eggs and egg case is little more than a memory-recall test; whereas to have the pupils predict the diet of a particular bird (whether it is a herbivore or carnivore or omnivore) given the kind and shape of beak the bird has is a test of thinking at a higher level and is thus preferable.

However, this is not to say that testing on facts is not to be encouraged at all. There is a place for questions that require factual answers; in these instances, the facts tested should be associated with the understanding of a science concept, process or method. A case in point is a question requiring pupils to state the characteristic features of the class of animals conceptualised as insects – three body parts (head, thorax and abdomen) and six legs.

Question 3.2 is regarded as an item that tests an esoteric fact because the answer involved is not common knowledge. An examination paper that contains too
many questions that test either trivial or esoteric facts is a concern because it could give the wrong signal to pupils, especially the conscientious ones, that learning science and doing well in science is simply about memorising a huge number of such facts. This would impede the development of creative and higher order thinking in them as their minds would be less free to roam and create, but instead would be distracted and cluttered with facts which should be more appropriately stored in places such as internet servers, books and CD ROMs. Such kinds of questions are counterproductive in terms of the TSLN drive.

**Example Question 3.1**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The eggs hatch into nymphs.</td>
</tr>
<tr>
<td>2</td>
<td>The eggs are protected by an egg case.</td>
</tr>
<tr>
<td>3</td>
<td>The egg case is hard, brown and purse-like.</td>
</tr>
<tr>
<td>4</td>
<td>There are only 4 eggs in the egg case.</td>
</tr>
</tbody>
</table>

**Example Question 3.2**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>goldfish</td>
</tr>
<tr>
<td>2</td>
<td>mackerel</td>
</tr>
<tr>
<td>3</td>
<td>swordfish</td>
</tr>
<tr>
<td>4</td>
<td>tilapia</td>
</tr>
</tbody>
</table>

**Comment**

As noted by Boo and Tan (2003) questions such as 3.1 and 3.2 do not test important learning outcomes in science. This category of problematic questions is best addressed through detailed vetting and stringent quality control throughout the examination setting process. Setters and veters of examination questions should be provided with clear guidelines on the expected question type profile for each examination paper.

**Category 4:**

**Questions with rigid and/or narrow mark schemes**

In the case of the open-ended or supply type items, it is common to find mark schemes that contain a narrow range of acceptable answers, and where an item has two (or more related) parts (see question 4.2 below), the marking rule adopted is that if the answer to the first part of the item does not match the answer given in the mark scheme, then the answer to the subsequent parts of the item will not need to be marked. In other cases, mark schemes could be too rigid in terms of requiring a specific (but not necessarily key) word or term to be given by pupils in order to earn marks or credit (see question 4.1 below). It is also common to find mark schemes that explicitly state a penalty for incorrect spelling or grammar.
Example Question 4.1

Susan and Michelle place a healthy plant in a beaker of water as shown in the diagram below. (Diagram of a plant with part of its stem and all of its roots immersed in a beaker of water.) They then cover the surface of the water completely with a layer of oil. The water level in the beaker was checked daily.

(a) After a few days, the water level in the beaker was lower. Give a reason for this. (2m)

(b) Why did they put a layer of oil on the surface of the water? (1m)

Comment

In this question, the question setter’s mark scheme gives examples of acceptable as well as unacceptable or incomplete answers. For part (a) the mark scheme specifies that examples of partially correct/incomplete/wrong answers are “Plants take in water” and “Water is lost”, while correct or complete answers are “The roots/root hairs of the plant take in/absorb the water” and “Transpiration (water loss) can cause an uptake of water”.

For part (b) the mark scheme gives the model answer as “The layer of oil prevents the water from evaporating”, and gives the following examples of partial/incomplete/wrong answers: “the oil stops air from entering”; “the oil prevents air from dissolving”; “water cannot escape; and water becomes water vapour”.

It seems to this author that mark scheme is rather rigid and narrowly focused on the key word “roots/root hairs” in part (a). This is unfair and counter-productive to the TSLN vision because it encourages rigid thinking and inaccurate focus. This author in fact thinks that the answer “plants take in water” which is regarded by the setter as only partially correct is in fact the more accurate or superior answer in the context of the question. The following analogy should clarify this: Suppose this author were to slap you, the reader, with my hand. Would you say “Your hand (a part of you) slapped me or you (the whole organism) slapped me?”
Example Question 4.2

Above are drawings of two Morning Glory plants, S and T of the same species. One was growing in a deserted field and the other was growing in the shade.

(a) Which of the above plants was likely to be found in the deserted field? (1m)
(b) Explain your answer in (a) (1m)

Comment

In the above question, the setter’s model answer for part (a) and (b) respectively were “S” and “Bigger leaves/flowers, must have received more sunlight”. However a pupil who is thinking divergently could produce different answers such as “T”, for part (a); and “deserted field has barren or poor soils hence smaller leaves/flowers” for part (b). If the marker follows the marking rule that part (a) of the answer does not match the model answer, then part (b) needs not be marked, then, the pupil’s answer as given would not be given any credit. Such pupils could become confused or even de-motivated in learning science.

A better approach to marking such open-ended questions is to have a broader range of acceptable answers, and to mark pupils’ answers based on the quality of their thinking, and to give due credit even in cases whether their answers are completely different to those supplied by the setter.

How Assessment Can Be More Aligned With The TSLN Vision

The conclusion arising from the scrutiny of these 100 assessment papers is that while there are questions that are well-set and effectively assess higher order thinking skills in pupils, there are also many other questions which are either disadvantageous to pupils who are capable of thinking laterally or creatively or deeply or which are counterproductive to our efforts towards achieving the TSLN vision.

The challenges involved in using multiple-choice questions to test pupils’ understanding of science concepts and processes and at the same time cater to the vision of TSLN have been highlighted. On the one hand MCQ items need to be very tightly bounded in order to maintain scientific correctness and uniqueness of the
answer key. On the other hand, the answer could be made obvious with the additional qualification, even to pupils who may not know the science content involved.

One important argument in favour of MCQ (in addition to the wider scope of coverage of topics to be assessed) is the capability to employ automatic marking which saves considerable marking time for the teacher as compared to open-ended questions. However, given the increased care and detailed thinking about possible alternate views required from the teacher, this is a debatable advantage.

It might be argued that in order to achieve the vision of TSLN, there is perhaps a need for change in our current mode of assessment in primary science. The change could involve introducing two-tier questions, in place of traditional MCQs, in the paper-pencil tests. This means that pupils are asked to respond to test items at two levels; hence the term “two-tier questions” are used in the literature (Treagust, 1988; Tyson & Bucat, 1995). In these two-tier questions, at the first tier, pupils select what they think is the most appropriate response out of four given options. At the second tier, they write a justification for their choice of a particular option. This would ensure that the perspectives of pupils who are competent in creative or higher order thinking are taken into consideration during marking. It would also mean that a longer time would be spent in the marking of such two-tier test items. However, this additional “cost” in time taken for marking could be justified in terms of the gain that results from its positive contribution in taking into account the divergent or higher order thinking of pupils.

The development of creative, innovative and higher order thinking is challenging and needs to be supported by appropriate assessment modes, and cannot be measured by paper-pencil techniques alone. This means that yet another way of ensuring that formal assessment is more aligned with the TSLN vision is to reduce the weighting given to paper-pencil tests by including more authentic kinds of assessment such as project work ( investigative as well as modelling, model-making), performance assessment involving hands-on manipulative tasks (Ang, Boo & Toh, 2003) or teacher-assessment of pupils’ ability to apply science concepts, skills and processes in the context of performance-based investigations (similar to the School Practical Assessment or SPA that has been introduced at the secondary and junior college/pre-university level). Such changes to primary school science assessment are worth considering in order to meet the objectives of TSLN since project work and performance-based investigations are more suitable for assessing higher order thinking skills such as identifying problems, generating hypotheses and designing and carrying out strategies to solve problems (Ministry of Education, 2000 p 3-2).
References


