

---

Title	Primary Science Assessment in the Context of Thinking Schools Learning Nation Vision.
Author(s)	Boo Hong Kwen
Source	<i>Teaching and Learning</i> , 24(2), 131-140
Published by	Institute of Education (Singapore)

---

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.

# Primary Science Assessment in the Context of Thinking Schools Learning Nation Vision

Boo Hong Kwen

## Abstract

*This paper arises out of a concern that since 1997 schools have been paying special attention to the development of creativity, lateral thinking and problem solving skills in our school going population; and that pupils have been taught to be deeper, more divergent, and more creative in thinking while at the same time there seems to be inadequate change in the formal assessment modes that are practised in schools. This paper highlights some instances where formal assessment in primary science does not match the vision of achieving "Thinking Schools, Learning Nation".*

## Introduction

In 1997, a nation-wide initiative to promote creativity, lateral thinking and problem solving skills was launched in Singapore under the slogan "Thinking Schools Learning Nation (TSLN)" (Goh, 1997). Since then schools have been paying special attention to the development of creativity, lateral thinking and problem solving skills in our students. Teachers across the various disciplines have been trained to teach such thinking skills explicitly, as well as indirectly, through infusion into their lessons (Science included). Pupils have been taught to be deeper, more divergent, and more creative in thinking. But has school assessment developed to keep in step with the TSLN drive? In this article, the focus is on examining this question with respect to formal school science assessment at the primary level.

The word creativity has a wide range of meanings. However, it generally means the production of something new (de Bono, 1986). There is some overlap between "creativity" and "lateral thinking". Lateral thinking, which is specifically concerned with the generation of new perceptions and new ideas, overlaps with creativity since both are concerned with producing something new, such as new ways of viewing a phenomenon or alternative ways of asking questions. In the context of education, the term "creativity" is often used as a value judgement of a result while lateral thinking is a process. It can be said that the ability

to think laterally is involved in creativity. In this paper, the term “higher order thinking” will be used to include creative or lateral thinking as well as depth in thinking.

## **Problematic Test Items**

### ***Background***

In these past three years, the author has acted as a consultant to several primary schools in Singapore to assist teachers in the vetting of science examination papers for levels Primary 3 (P3) to Primary 6 (P6). For the P5 and P6 levels, only the EM1/2 Science papers are included in this analysis since pupils in the EM3 stream do not take Science as an examination subject at the Primary School Leaving Examination (PSLE). (To maximise their potential, pupils are formally streamed according to their learning ability at the end of P4. For P5 and P6, pupils are placed in one of three language streams, namely EM1, EM2 and EM3, according to their abilities. Pupils in the EM1 and EM2 streams do English Language, Mother Tongue, Mathematics and Science. EM1 pupils may do Higher Malay/Higher Chinese/Higher Tamil as their Mother Tongue. Pupils in the EM3 stream do Foundation English Language, basic Mother Tongue and Foundation Mathematics and are not tested in Science at the PSLE.)

The procedure adopted was for the teachers to provide final draft forms of the examination papers for critical vetting with the aim of providing feedback to the teachers concerning the strengths and weaknesses of the test items. During the period approximately 100 papers were examined. Each paper typically comprising 30 multiple choice questions (MCQs) and 16 open-ended supply type or free response questions.

The process of vetting involved detailed examination of approximately 4600 test items *vis-à-vis* the 13 principles for test item crafting articulated and elaborated in Boo and Tan (2003). Further examination of the test items that violated at least one of the 13 principles was then undertaken to determine if there were particular questions that might disadvantage pupils who are more divergent or deeper in their thinking or which were not in alignment with the objectives of TSLN. From this further examination, four categories of flawed items were identified that were not in alignment with the objectives of TSLN.

It should be noted that while there were test items which were flawed in these four particular ways, there were also well-crafted test items which effectively test pupils' higher order thinking skills and acquisition of science concepts and which were also fully aligned with the TSLN objectives. The potential problems arising in the four identified categories are discussed below, together with some approaches that would help alleviate them.

## Four Categories of Items that Are Not Supportive of Higher Order Thinking

### ***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 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 suggests that for such pupils, there is an additional obstacle to performing well on conventional paper–pencil test items since they are likely to have alternative perceptions or interpretations of a question. This is particularly severe in the case of 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 a 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**

Which one of the following animals should *not* be in the same group as the others?

- 1 hen
- 2 goat
- 3 tiger
- 4 rabbit

#### **Comment**

In this question, the setter's answer key is option 1 "hen". 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 "hen" is clearly the odd-one out, the only bird among the mammals. 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 tiger is the only carnivore.
- Option 4 could be the answer key since rabbit is the only animal that burrows and bears its young underground.

For addressing the problem of perceptual mismatch between the question setter and the creative pupils, it would be worth considering the inclusion of an open-ended section following every multiple-choice question. 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 creative pupils 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 can 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.

Another approach 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 more obvious and hence easily selected by pupils, as could be seen from the improved version of Question 1.1 shown below.

In this improved version of the question, the basis of grouping the animals (i.e. the type of body covering) has been specified, and there is now clearly one and only one acceptable answer key (option 1).

**A suggested improved version of Question 1.1**

Based on type of body covering, which one of the following animals should *not* be in the same group as the others?

- 1 hen
- 2 goat
- 3 tiger
- 4 rabbit

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

**Category 2: Items that Are Incorrectly Focused  
on Low-Level Recall or Extraction of Information**

Included in this category are items where the design of an investigation is given and the pupils are required to suggest the aim of the investigation. This type of test item simply encourages recall of the aim of an investigation that has been discussed or carried out in class and does not encourage higher order thinking. This is counterproductive to the TSLN impetus.

**Example Question 2.1<sup>1</sup>**

Joan carried out an experiment in her living room with 4 stalks of flowers cut from the same plant. She recorded the results of her experiment in a table as shown below.

Vase	A	B	C	D
Number of stalks of flowers	1	1	1	1
Type of chemicals	P	Q	R	S
Amount of water (ml)	500	500	500	500
Number of days flowers stayed fresh	7	5	6	3

- (a) What was Joan trying to find out from her experiment? (1m)  
 (b) Besides using similar vases in size and quality, state 2 other variables which she must keep the same for the experiment. (2m)

<sup>1</sup>The example given is reproduced directly from one of the examination papers. This question, as originally conceived, is also weak in terms of the design of the investigation presented. This is discussed in more detail in Boo and Tan (2002).

**Comment**

That the required answer to part (a) of "testing the relationship between the type of chemical and the number of days the flower stalks stay fresh" is obvious from the table without any thinking of the science content involved since these are the only two values that vary. The answer to part (b) of "same amount of water" and "same number of flower stalks" is even more obvious from the table, and does not require any consideration of the science content or processes involved at all.

There would be far greater value in presenting the relationship to be tested and requiring the pupils to provide a detailed experimental design. This approach (which would subsume part (b)) would 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 an experimental control.

### **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 higher order thinking in pupils. Question 3.1 which simply requires pupils to recall how many teeth a healthy adult should have (a fact which is regarded as trivial because it is regarded as common knowledge) is little more than a memory-recall test; whereas to have the pupils predict the diet of a particular animal (whether it is a herbivore or carnivore or omnivore) given the kind and arrangement of teeth the animal 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 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 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 kind of questions are counterproductive in terms of the TSLN drive.

#### **Question 3.1**

How many teeth does a healthy adult have?

- 1 20
- 2 24
- 3 28
- 4 32

#### **Question 3.2**

Which of the following has leaves that fold up at night?

- 1 angkana
- 2 flame of the forest
- 3 orchid
- 4 rain tree

### **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 items is best addressed through detailed vetting and stringent quality control throughout the examination setting process. Setters and vetters 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 Narrowly Based Marking Schemes**

In the case of the open-ended or supply type items, it is common to find marking schemes that contain a narrow range of acceptable answers, and where an item has two (or more related) parts (see Question 4.1), the marking rule adopted is that if the answer to the first part of the item does not match the answer given in the marking scheme, then the answer to the subsequent parts of the item will not need to be marked. In other cases, marking schemes may 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. It is also common to find marking schemes that explicitly state a penalty for incorrect spelling or grammar.

The concern here is that pupils who are capable of higher order thinking see alternatives which are often not seen by the setter, and if the marking schemes are followed rigidly, these pupils end up giving quite correct answers to the question as posed but which is/are the wrong answer(s) as per the question setter's marking schemes. Such pupils may then become confused or even de-motivated in learning science. Here, again, such questions work against our vision of achieving TSLN.

#### **Example Question 4.1**

Given a diagram showing a solar powered car and a hot-air balloon and the following table:

Mode of transport	Energy source
Hot-air balloon	Burns fuel
Solar-powered car	Sun's energy

- (a) Which mode of transport is environmentally friendly? (1m)  
 (b) Explain your answer in (a). (1m)

Answer given on setter's marking scheme:

- (a) solar-powered car  
 (b) Sunlight/solar battery/cell/solar energy does not cause pollution/  
 conserves natural resources

### Comment

Such a narrow marking scheme is disadvantageous to pupils who are thinking divergently or deeply. For example in this question, while the setter is thinking of the renewable or the clean nature of solar energy, pupils who are well-versed in science concepts and capable of lateral or deep thinking may come out with an answer that is different from that of the setter. Such pupils may argue that the hot-air balloon is more environmentally friendly than the solar-powered car based on two reasons: first, it does not require as much energy to manufacture as the solar-powered car, and second, it does not require the clearing of forests and the construction of roads. There is yet another reason that could be given in favour of the hot-air balloon – that it could use hydrogen as a fuel, which would then produce water as the only by-product (which is non-polluting). In such a case, if the setter adopts the marking rule that if part (a) of a pupil's answer does not match the answer given in the marking scheme, then part (b) will not be marked, pupils who give answers such as those described above, which are different from the setter's, but which are valid, will be awarded "zero" marks.

A more acceptable marking scheme for Question 4.1 is as follows: Answer to part (a) could be either solar-powered car or hot-air balloon depending on quality and soundness of reasoning given in part (b). Award 2 marks for parts (a) and (b) if at least one plausible reason exemplified in the table below is given.

Part (a)	Solar-powered car	Hot-air balloon
Part (b)	Sunlight/solar battery/cell/ solar energy does not cause pollution/conserves natural resources	Does not require as much energy to manufacture as the solar-powered car/uses hydrogen as fuel hence very clean/does not require the clearing of forests and the construction of roads

During the marking process, markers should be open and be prepared to modify marking schemes as and when valid answers (which have not been anticipated in the marking schemes) are given by pupils.

Also, whilst good spelling and grammar should be encouraged in all subjects and not just the formal language lessons, there is little justification for penalising incorrect spelling if the words are recognisable or penalising grammatical errors if the answers make sense in terms of the science content. (Cambridge examiners for "A" and "O" level Science subjects do not penalise students for incorrect spelling if words are recognisable nor for grammatical errors.) Correct spelling and good grammar can and should be encouraged in science work done through the school year. But in formal tests, such as end-of-year summative examinations that will lead to pupils' formal ranking or streaming, the marking of free response or supply type items should only be on the correctness of the science concepts and the quality of reasoning or thinking that are revealed in the answers.

## **Discussion – How Assessment Can Be More Aligned with the TSLN Vision**

The conclusion arising from the scrutiny of these 100 primary assessment papers is that while there are questions that are well-set, and effectively assess higher order thinking 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.

Whilst some specific suggestions have been provided to address the identified categories of problematic questions, it is worth stepping back to question the current approach adopted in question setting.

The challenges involved in using MCQs 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, this magnifies another problem of MCQ, and that is, 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 of employing 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 alternative views, which is 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 may involve introducing two-tier questions instead of the traditional MCQ. Another approach is to reduce the weighting given to paper-pencil tests by including more authentic kinds of assessment such as project work assignments and teacher-assessment of pupils' ability to use 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 (MOE, 2000, p. 3-2).

## References

- Boo, H. K. & Tan, K. C. D. (2003). Crafting and marking paper-pencil test items. In K. C. Yap, K. A. Toh, & N. K. Goh (Eds), *Primary Science Teaching: A Sourcebook for the Primary School Teacher* (pp. 146–159). Singapore: Pearson Prentice Hall.
- Boo, H. K. & Tan, K. C. D. (2002). Enhancing the standard of project work in primary science. *ASCD Review*, **10**(3), 15–19.
- de Bono, E. (1986). *CoRT Thinking: Creativity (Teacher's Notes)* (2nd edn) (p. 7). U.K.: Pergamon Press.
- Goh, C. T. (1997). *Shaping our future: Thinking Schools, Learning Nation*. Speech at the Opening of the 7th International Conference on Thinking on Monday, 2 June 1997. Available: <http://www1.moe.edu.sg/speeches/1997/020697.htm>
- Ministry of Education (2000). *The Science Department Handbook (Secondary and PreUniversity)*. Singapore.
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, **10**(2), 159–169.
- Tyson, L. M. & Bucat, R. B. (1995). Chemical equilibrium: Using a two-tier test to examine students' understanding. *Western Australia Science Education Association Annual Conference*, 1995.