
Title	Is this really what you mean? Problems of perception in science multiple choice questions
Author(s)	Boo Hong Kwen
Source	<i>NZARE - AARE Joint Conference, Auckland, Australia, 29 November to 3 December 2003</i>

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.

IS THIS REALLY WHAT YOU MEAN? PROBLEMS OF PERCEPTION IN SCIENCE MULTIPLE CHOICE QUESTIONS

Boo Hong Kwen

National Institute of Education
Nanyang Technological University,
Singapore

ABSTRACT

Assessment is an integral part of the teaching-learning process, and typically consumes a significant portion of both the teachers' and pupils' time and energy. The background of this paper concerns the fact that multiple choice questions (MCQs) are used rather extensively in the assessment of science learning outcomes from year three through to year six at the primary school level, and the fact that some of these questions involve a mismatch in perception between the test setter(s) and the test takers. This discussion paper is based on an investigation into 100 sets of primary school science examination papers from 25 different schools. It discusses specific examples of perceptual mismatch between the question setter(s) and the test takers. It also highlights the concern that too much reliance on the use of traditional MCQs could seriously disadvantage pupils who are more divergent or creative or deeper in thinking, as well as impede the development of creativity in pupils. Suggestions for addressing some of the concerns raised are also included in the paper.

Key words: science assessment, primary science, and multiple-choice questions

BACKGROUND AND INTRODUCTION

Through years 2001, 2002 and 2003, the author conducted some school-based workshops and training courses on primary science assessment. She also acted as a consultant and helped to vet science examination papers in a number of schools. Prior to the conducting of the workshops and training courses, participants/schools were requested to send in their schools' science examination papers (from primary 3 through primary 6) used in the previous year's examinations.

During the period, about 100 sets of examination papers were scrutinised, each set of papers typically comprising 30 multiple choice questions (MCQs) carrying 60% of the marks, and 16 open-ended supply type or free response questions, carrying the remaining 40% of the marks. It could be seen that a higher weighting, both in terms of number of questions and of marks allocated, has been allocated to the MCQ section of the papers.

It should be noted that while there were test items that were flawed, there were also many well-crafted test items that effectively test pupils' creative and higher order thinking skills, as well as their acquisition of science concepts. However, what is reported the following section is a representative sample of the MCQs, which are problematic in the sense that they have potentials for involving mismatch in perception between the question setter(s) and pupils taking the test, and in particular, the more creative pupils. With the significant proportion of marks allocated to MCQs, this means that creative pupils could be disadvantaged in the tests/examinations. The critique and discussion of the open-ended questions in the examination papers are not included in this paper, but can be found in Boo (2002, 2003a, 2003b, and 2003c).

QUESTIONS THAT INVOLVE PERCEPTUAL MISMATCH BETWEEN QUESTION SETTER AND PUPILS

In some of the test items, there are potentials for mismatch in perception between the question-setter and the test-takers. 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. However, this problem of mismatch in perception of the question posed is more pronounced where pupils who are creative or are capable of higher order thinking are concerned. This is because these pupils tend to see issues, alternatives and ambiguity in test items which 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 the 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 automatically, i.e. by a computer. In some MCQs (such as questions 1 and 2 below), pupils who are thinking laterally or deeply, find situations in which all or none of the options are correct while pupils who are not as capable in their thinking or who know the material only in superficially, or in the conventional way, could simply select the most appropriate (often most obvious) answer and get marked correct.

Example Question 1

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

- 1 cheetah
- 2 chicken
- 3 goat
- 4 rabbit

Comment on Question 1

In this question, the setter's answer key is option 2 "chicken". 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 2 "chicken" 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 1 could be the answer key since "cheetah" is only carnivore.
- Option 3 could be the answer key since "goat" is the only animal with horns.
- Option 4 could be the answer key since "rabbit" is the only animal that burrows and bears its young underground.

Example Question 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

Comment on Question 2

In this question, unlike the previous one, the setter has specified the basis of classification in terms of movement of the animals. The model answer given 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 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.

Example Question 3

The diagram below shows the human skeleton. (A diagram of the human skeleton is given, with ball and socket joints at one shoulder and hip are labeled as A and C respectively, with hinge joints at the elbow and knee labeled as B and D respectively.) Which of the joints indicated, allow the greatest freedom of movement?

- 1 A and C
- 2 B and D
- 3 A, B and C
- 4 B, C and D

Comment on Question 3

In this question, the setter’s answer key is option 1. Here the setter is probably thinking of the fact that the ball and socket joints at the shoulder and at the hip are the joints that individually allow the greatest freedom of movement. However, it is possible that some pupils could be thinking mathematically, in terms of the sum of the movements allowed by the joints mentioned in each option and hence, choose option 3, rather than option 1 as the answer, because option 3 includes the two ball and socket joints (A and C) plus a hinge joint, B.

Example Question 4

Arrange these animals from the slowest to the fastest.

- | | |
|---|----------|
| A | cat |
| B | cheetah |
| C | tortoise |
| D | antelope |
1. A, B, C, D
 2. C, A, B, D
 3. D, B, C, A
 4. C, A, D, B

Comment on Question 4

In this question, the setter gives the answer as option 4 “tortoise, cat, antelope, cheetah”, perhaps thinking of movement over short distances, or having overlooked completely the question of short or long distances. However, some pupils, thinking of movement across long distances, could choose option 2 as the answer, since over long distances the antelope is a faster runner than the cheetah. Yet another group of pupils could be perplexed by an apparent overlap between alternatives A and B since the cheetah is a member of the cat family.

Example Question 5

Four solid balls, A, B, C and D, of the same size, but made of different materials, are arranged in order, starting with the heaviest. Which one of the following is the correct order?

- A wooden ball
- B plastic ball
- C metal ball
- D paper ball

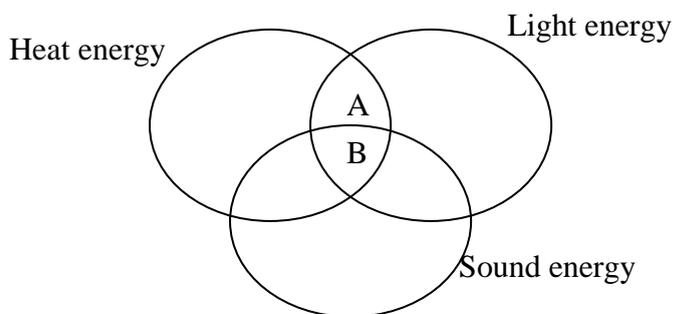
- 1 B, C, A, D
- 2 B, A, D, C
- 3 C, B, A, D
- 4 C, A, B, D

Comment on Question 5

The model answer to this item is given as option 4. However, some pupils could be reasoning correctly and yet give an answer which is different to that intended, such as option 3. The problem here is that there are various kinds of wood, plastic and metal. The pupils may have come across plastics which are denser than wood. Because the specific kinds of plastic and wood have not been specified, it is possible that the pupils could have in mind a kind of plastic which is denser than a certain kind of wood, while the question setter has in mind yet another kind of plastic which is less dense than a certain kind of wood. This is because plastics have a range of densities (and properties), ranging from those less dense than water and certain types of wood, to those which are denser than water and certain other types of wood! Wood too has a range of densities, though not as varied as plastics. E.g., Pine wood and balsa wood are less dense compared to many kinds of plastics but denser than some other plastics such as polystyrene!

Example Question 6

Study the Venn diagram below where each circle represents a different kind of energy.



Which of the following activities can A and B possibly be?

	A	B
1	Striking a match	Turning on television
2	Bouncing a ball	Having a hot shower
3	Turning on radio	Switching on a light bulb
4	Lighting a candle	Cooking popcorns

Comment on Question 6

Here the question stem is vague or unclear. It should have included some statement to the effect that it concerns the different kinds of energy involved in two activities, A and B. The model answer given is option 1, which suggests that the question setter is thinking of the heat and light energies that are produced when the match is struck, and the light, sound and heat energies that are produced when the television is turned on. However some pupils could choose option 4 as the correct answer, since

lighting a candle involves light and heat energy, and cooking popcorns in a pressure cooker over a gas stove produces heat, light and sound energies. The problem here is that the question setter is perhaps thinking of cooking popcorn in some other device and/method where no popping sound is involved.

There could also be some pupils who traced the series of changes a step farther back than the question setter and are perplexed in the question and could not find a suitable answer - because these pupils are thinking of the fact that movement energy must be involved in striking a match and hence option 1 (as well as option 4 – since lighting a candle generally involves striking a match) could not be the answer.

Example Question 7

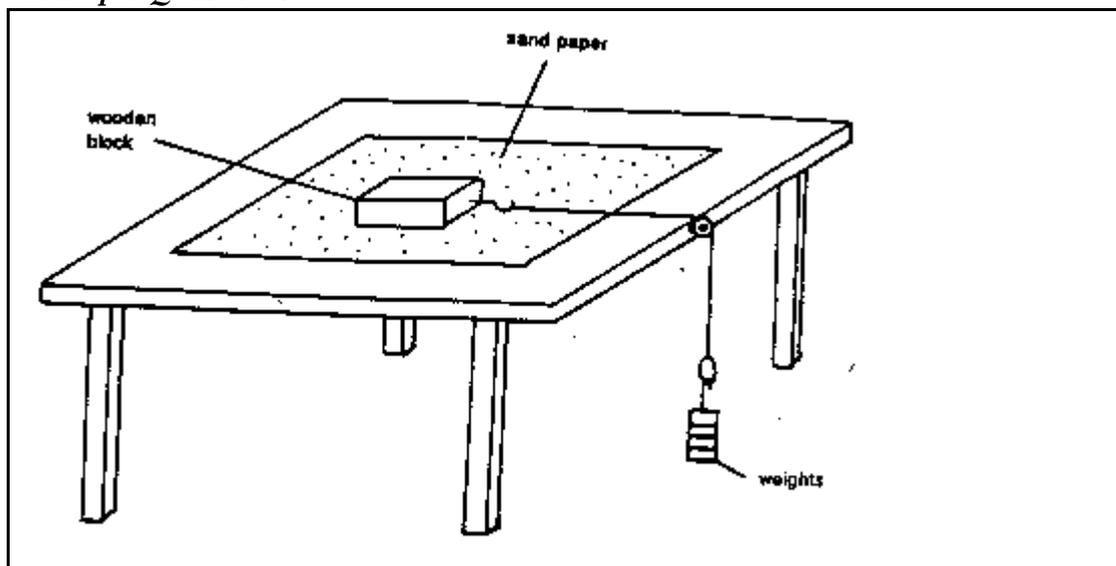
Which of the following gases is produced during decomposition?

- 1 oxygen
- 2 nitrogen
- 3 helium
- 4 carbon dioxide

Comment on Question 7

Here again, the question stem is vague. The problem with this question, where the model answer given is option 4, is that the setter is thinking only of decomposition of organic matter and assumes that the test takers will be thinking of the same process or phenomenon. However, some pupils who may have read of, or have done experiments involving, decomposition of inorganic matter/substances such as hydrogen peroxide (which decomposes spontaneously to produce water and oxygen) may choose option 1 as the correct answer. There are also probably some other inorganic substances which undergo decomposition to yield nitrogen.

Example Question 8



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.

Type of surface	Sandpaper	Carpet	wood	glass
Force required/gf	60	90	?	35

What would be the force needed to move the wooden block over the wooden surface?

- (1) 30gf (2) 45gf (3) 65gf (4) 80gf

Comment on Question 8

This question involves non-living things which are more predictable than living things. However, 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. The question above as it stands does not exemplify good experimental procedure because there is no indication 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 examination takers would have in their minds exactly the same kinds 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!

Example Question 9

Classify the following 5 objects into the appropriate boxes in the table below. (5 m)

Fishing rod, Flag pole, Pencil sharpener, Door knob, Egg beater

Inclined plane	Lever	Pulley	Wheel and axle

Comment on Question 9

Here the problem is the lack of diagrams or pictures to show the specific kinds of objects that the question setter has in mind. Each of the five named objects has variations which utilize different principles or types of simple machines. For example, the model answer has door knob, egg beater and the pencil sharpener grouped under "wheel and axle", while the flag pole is placed under "pulley" and fishing rod under "lever". The problem here is that pupils could have in their minds different sorts of pencil sharpeners, door knobs, egg beaters, flag poles and fishing rods compared to those that the question setter has in mind. For example, while the setter is thinking of a kind sophisticated device which makes use of the wheel and axle principle, some pupils could be thinking of the simplest kind of pencil sharpener which uses a sharp metal blade which perhaps could be classed as an inclined plane. In the same way while the teacher is thinking of a rather complicated device which utilizes gears as well as the wheel and axle principle, the pupils could be thinking of a different kind of egg beater such as a simple fork!

SUGGESTIONS ON ADDRESSING THE PROBLEM OF POSSIBLE PERCEPTUAL MISMATCH BETWEEN THE QUESTION- SETTER AND THE TEST-TAKERS

The conclusion arising from the scrutiny of these 100 sets of examination papers is that while there are questions that are well-crafted and could effectively assess higher order thinking in pupils, there are also other questions that are disadvantageous to pupils who are capable of creative and higher order thinking.

Traditional MCQs such as the first eight examples discussed, in which there is a question stem (which poses a problem in the form of either a direct question or an incomplete statement) plus four alternatives or options (which are either possible answers to the direct question or completions to the incomplete statement) have been widely used because of the wide coverage of topics afforded by them as well as the reliability and ease of marking (usually by automatic or computerised means).

However, in a context where pupils are increasingly being trained to be creative and critical thinkers, as well as good problem solvers, it is becoming increasingly challenging for setters of science assessment items to be able to craft such traditional MCQs where there is one, and only one, unique

answer key. This is because science, which can be regarded as man's systematic inquiry of the natural world, more than any other subject, has a rather wide scope of coverage. The result is that there are potential areas for alternative interpretations and perceptions of words, diagrams and illustrations presented in questions.

One way of addressing the problem is to be as specific and explicit as possible in the framing of both the stem and the options. This is illustrated in the following improved version of the original question 1 discussed in the previous section.

Improved version of Question 1

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

- 1 cheetah
- 2 chicken
- 3 goat
- 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 chicken has feathers as a body covering while the other three animals have hair.

Test setters who find such detailed specification or refinement unacceptable in the sense that it makes the test item too easy or unchallenging for pupils may consider yet another way of addressing the problem. This second approach is to introduce two-tier questions, in place of the traditional MCQ. 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 can be justified in terms of the gain that results from its positive contribution in taking into account the creative or higher order thinking ability of pupils.

One important argument in favour of MCQs (in addition to the wide coverage of topics) is that they allow for automatic or computerised 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 and interpretations required from the teacher, this is a debatable advantage. The two-tier questions could be a good compromise, a possible solution in addressing the need to have wide coverage of the topics on the test as well as catering to alternate perspectives among creative pupils.

On a broader level, it could be argued that comprehensive assessment of learning outcomes in science cannot be achieved through the use of paper-pencil tests alone. Science learning outcomes include content (concepts, models, and principles), skills and processes as well as attitudes and values, and need to be supported by appropriate assessment modes, and cannot be measured solely through paper-pencil techniques. From this perspective, another possible solution 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 is being phased in at the secondary and junior college/pre-university level).

REFERENCES

- Ang, K.C., Boo, H.K. & Toh, K.A. (2003). Performance of primary 4 students at performance-based assessment tasks. *Educational Research Association Singapore (ERAS) Conference*, 19-21 November 2003, Singapore.
- Boo, H.K. (2002). Primary science assessment and its contribution to the nurturing of creativity. *Australian Association for Research in Education (AARE) Conference*, 1-5 December, 2002, Brisbane, Australia.
- Boo, H.K. (2003a). The role of school science assessment in enhancing students' creativity. *Asia Pacific Conference on Education*, 2-4 June 2003, Singapore.
- Boo, H.K. (2003b). Primary science assessment in the context of "Thinking Schools Learning Nation (TSLN)" vision. *Teaching and Learning*, December (2003) Issue (in press)
- Boo, H.K. (2003c). Does formal assessment in primary science support the "Thinking Schools Learning Nation (TSLN)" vision? *Educational Research Association Singapore (ERAS) Conference*, 19-21 November 2003, 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.