

Characteristics of Multiple Choice Questions Intrinsic to Their Format

Taiichi SAITO, Seizaburo ARITA* and Ikuo NASU**

*Department of Pharmacology, *Department of Mathematics,
Kawasaki Medical School, Kurashiki 701-01, Japan*

***Department of Dental Public Health, Nihon University
School of Dentistry at Matsudo, Matsudo 271, Japan*

Accepted for publication on July 15, 1988

ABSTRACT. Multiple choice questions (MCQs) are not suitable for testing an examinee's knowledge. Examiners cannot distinguish whether questions have been answered by random guessing or not. In many types of MCQs, knowledge of only two items (a correct terminal) out of four or five is sufficient to answer correctly. If an examinee can identify a correct terminal, he can answer a question correctly even if all other items are blank. In the multiple completion type problem (K-type) and its modifications, one to five correct terminals exist depending on answer codes. Therefore, even if an examiner presents difficult material in an item, an examinee could receive points by locating another correct terminal. On the other hand, even if an examinee knows all three correct items appearing in the answer code (e.g. type K, answer code "A"), he is still compelled to make a random guess and could select the wrong answer. Therefore the scores achieved by the examinees on such tests can not be said to truly reflect their actual knowledge of a subject. MCQs can, however, be used for self-assessment.

Key words : multiple choice questions — correct terminal —
distorted evaluation of achievements

There have been many arguments for and against the use of multiple choice question (MCQ) in medical schools and in licensing examinations. Among articles concerning this matter, those of Pickering¹⁾ and Anderson^{2,3)} have been comprehensive. Anderson favors employing MCQs to evaluate the achievement of medical students and licensing candidates. Pickering critically discusses the effect of using MCQs on medical education and the attitude of medical students. Eichna⁴⁾ has correctly observed that "Since examinations today glorify facts and involve little thinking or problem solving, students cram facts and give minimal thought to everything else. The predominant influence has been the objective type of examination given by the National Board of Medical Examiners. With few exceptions, Board-type examinations have become the usual examinations in medical school". This is also the case in Japan. MCQs have monopolized the National Board Examination for Medical Practice (NBEx) in Japan since 1975. Pickering¹⁾ and Dudley⁵⁾ in dealing critically with MCQs believe that MCQs test only trivial (low-taxonomy) facts. Anderson³⁾ and Barker and Maatsch⁶⁾ agree though that these weaknesses of MCQs are not the fault of the test format. They believe that if test constructors are well trained

斎藤泰一, 有田清三郎, 那須郁夫

and suitable items are written, reliable tests can be provided. To date, however, certain characteristics intrinsic to the format of some types of MCQs have gone largely unmentioned. First let us consider the one/best response type of MCQ (one-from-five, or five-choice completion type, Type A of Hubbard and Clemens' classification⁷⁾). Let's look at two examples of this type of question.

Q1. Which city is located in Japan? Choose the true item.

- A 京 都
- B New York
- C London
- D Paris
- E Rome

Property 1. The knowledge of only one true item allows the examinee to get one point. Any Japanese, including those unable to read English, could easily answer this question correctly.

Property 2. This question could also be answered easily and correctly by anyone unable to read the characters appearing in A, because there must be a correct answer and it is clear that B-E are all false items. Now, let's go to the second example.

Q2. The stem of the question is the same as Q1.

- A 奈 良
- B 紐 育
- C London
- D Paris
- E Rome

Property 3. This time the examinee who does not know the two items A and B must resort to random guessing even if he knows that three of the items are false, because only one of the two unknown items can be true. If the difficulty level of only one false item out of four competes with that of the true item in this type of MCQ, the examinee will have to make a random guess.

It is possible to answer correctly if 1) only the true item is known or 2) the four false items are known. In Q1, therefore, item A was equivalent to items B~E. If any of the false items had been of the same difficulty as the true item and neither had been known, the examinee would not have been able to answer this question. Such a type of question must compel the examinee to make a random guess.

Property 4. The marks acquired are not proportional to the amount of the examinee's knowledge. If an examinee knows only the true item (20%), he or she can get a point, but other examinees who may know three false items (60%) will only be able to get a point by random guessing.

In multiple completion type (Type K⁷⁾ questions, an initial statement is followed by four completions (or items). The answer codes (combinations of true items) are as follows:

- A — Items 1, 2 and 3
- B — Items 1 and 3
- C — Items 2 and 4
- D — Only item 4
- E — Items 1, 2, 3 and 4

For this type of question, the examinee needs to know only two appropriate items (a correct terminal) to get the correct answer (Table 1). Suppose that

TABLE 1. Multiple true/false type (Type K) of MCQ and sufficient knowledge to solve the question

An initial statement (or stem) is followed by four completions (or items).
 The examinee must answer
 A — If items 1, 2 and 3 are correct.
 B — If items 1 and 3 are correct.
 C — If items 2 and 4 are correct.
 D — If only item 4 is correct.
 E — If item 1, 2, 3 and 4 are all correct.

Examinee can answer correctly by knowing only two items.

Answer Codes (Correct items)	Correct terminals (Combinations of least two items sufficient to answer correctly.)*	False terminals (Combinations of known items insufficient to answer correctly.)*
A (1, 2, 3)	2, 4	1, 2 1, 3 1, 4 2, 3 3, 4 1, 2, 3 1, 3, 4
B (1, 3)	1, 2 2, 3 2, 4	1, 3 1, 4 3, 4 1, 3, 4
C (2, 4)	1, 2 2, 3	1, 3 1, 4 2, 4 3, 4 1, 3, 4
D (4)	1, 2 2, 3 2, 4	1, 3 1, 4 3, 4 1, 3, 4
E (1, 2, 3, 4)	1, 4 3, 4	1, 2 1, 3 2, 3 2, 4 1, 2, 3

* Items not appearing in the answer codes are all false items.

the correct answer code is A. If the examinee knows a set of items, 2 and 4, he can find the correct answer. Other items (1 or 3) do not contribute to getting a point, and neither do the five remaining sets of two items and the two sets of three items shown in Table 1. Surprisingly enough, even if an examinee knew the three items written in answer code A, he could not answer this question without random guessing. Although the examiner clearly made this question with the intention of demanding correct knowledge of items 1, 2 and 3 from examinees, examinees who knew those items might not be able to answer correctly. On the other hand, examinees knowing only items 2 and 4 could find the right answer. In other words, an examinee with correct knowledge of 75% of the items could fail after random guessing, while one with correct knowledge of only 50% of the items (a correct terminal) could get a point. When the answer code is B, there are three correct terminals, (1, 2), (2, 3) and (2, 4), in which 1 and 3 are true, and 2 and 4 are false items. With any one of these combinations a correct answer can be achieved even when the alternative two items are blank. In such a case, even if the examiner makes item 3 difficult, an examinee may skip that item, answering easily using the other easier terminals (1, 2) or (2, 4). However, if only item 2 were difficult, all examinees would have to resort to random guessing to achieve the correct answer. This is true in all cases of type K question, unless the correct

TABLE 2. Modified multiple completion types used in the Japanese National Board Examination for Medical Practice

- I. An initial statement is followed by four completions (or items). Choose the correct code. (Modified K type)
- A. (1) (3) (4)
 - B. (1) (2)
 - C. (2) (3)
 - D. (4)
 - E. (1) ~ (4)
- II. An initial statement is followed by five completions (or items). Choose the correct code. (from A to E)
- | | |
|---|---|
| <p>1. Doublet type</p> <ul style="list-style-type: none"> A. (1) (2) B. (1) (5) C. (2) (3) D. (3) (4) E. (4) (5) | <p>2. Triplet type</p> <ul style="list-style-type: none"> A. (1) (2) (3) B. (1) (2) (5) C. (1) (4) (5) D. (2) (3) (4) E. (3) (4) (5) |
|---|---|

code is E. Therefore, examiners should not neglect these characteristics when setting the relevance, difficulty and minimal pass levels of MCQs.

For the past ten years only MCQs have been employed in the NBEx in Japan. There have been no essay type tests, oral examinations or practical tests used at all. About 45% of the questions on the MCQ test are of the one/best response type, while the remaining questions are of the modified multiple completion type. Three modifications of the multiple completion type are used (Table 2). When the NBEx was introduced in Japan, type K was omitted because items 1 and 3 are always linked together; i.e., if item 1 is true, item 3 is also true. Instead, a modified K type (Table 2, I) has been used. This modified K type has more correct terminals than the original type. This means that examinees have more routes by which they can find the correct answer. In addition, discrepancies between the routes which examiners intend examinees to take in finding answer and those which they actually do take may be greater. The other properties of this modification are the same as those of type K. There are two other types employed, a doublet type and a triplet type. The triplet type is a reverse type of the doublet type (Table 2, II). In both types there are four sets of correct terminals for each answer code (A~E). A correct terminal consists of only two items and knowledge of these two items is sufficient to find the correct answer. In any case, when an examinee identifies any one of the correct terminals, he or she can answer correctly even if the alternative three items are blank. In the case of the triplet type, the middle item in the triplet ((2) in A, (1) in B,.....) does not contribute to identifying the correct answer. (Put five numbers in a circle, to better understand which item is in the middle.) Suppose that the correct answer code is A. We remove item (2), leaving items (1), (3), (4) and (5) in a circle serially (Fig. 1). The correct terminals are any two items standing side by side in the circle. In the case of the doublet type, after removing the second of three false items, then follow the same above-mentioned procedure. In both types, each answer code has four correct terminals. However, knowledge of two items on opposite sides in the circle (1 and 4; 3 and 5 in Fig. 1) will not contribute to finding the correct answer, because neither set is a correct

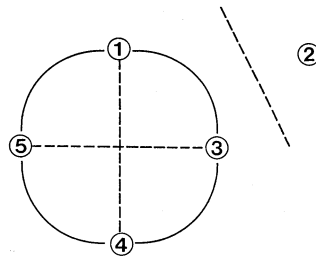


Fig. 1 How to find the correct terminals in Triplet types of MCQs. Remove the middle item of triplet items (correct answer: 1, 2, 3) and put the remaining items in a circle. The sets of two items standing side by side are the correct terminals: (1, 3), (3, 4), (4, 5) and (1, 5). This figure can be applied to both Triplet type (correct answer "A") and Doublet type (correct answer "E") MCQs.

terminal. However, if an examiner sets the difficulty level of one of the sets of opposite items ((1, 4) or (3, 5)) very high, an examinee would have to resort to random guessing even if the difficulty level of the alternative set was very low. In the doublet and triplet types of MCQ, the marks achieved also not proportional to the quantity of the examinee's knowledge.

In the one/best response type of MCQ, correct knowledge of only one true item or four false items will enable examinees to get a point. The five items are not absolutely independent and the one true item is equivalent to four false items. Therefore, the probability analysis so far applied to this type of MCQ does not always appear to be relevant.

In the multiple completion type and its modifications, an appropriate combination of only two items is sufficient to answer correctly. Remaining items do not contribute to identifying the correct answer. If one or more (depending on the format) items are very difficult in all of the above mentioned types, and examinee will be compelled to make a random guess. Owing to such properties, the marks achieved by the examinees may not be proportional to the quantity of their knowledge. They may be able to answer with partial knowledge of a subject.

There are several discrepancies which examiners have so far not been aware of. Examiners believe that all items are read by examinees, but many examinees can answer after reading only two items and the remaining items may be used, if at all, only for assurance. The intention of examiners may be dodged by examinees and the difficulty levels set by examiners may not be as high as the examiners think. In all types of MCQs, examiners cannot later identify which items were guessed randomly by examinees.

Most examiners believe that if the number of questions is sufficiently large, MCQs can be reliably used. This is only an illusion. In guessing heads or tails with a coin, it is true that the more trials, the less variation there will be. If coins are minted properly, the occurrence of heads is a random phenomenon. However, as mentioned above, an examinee with correct knowledge of only two items (a correct terminal) can always get a point, though another examinee with correct knowledge of three items might lose a point after a random guess. Therefore, the probability analysis should not be applied directly

to such a phenomenon. Moreover, coins are taken from a single population. The quality and difficulty level of questions or items, however, may be different in a test consisting of 10 questions and another consisting of 100 questions. Suppose that you make 10 questions and 100 questions at the same time after a single lecture. In the former case it is easier to collect 40~50 important items for inclusion in your test, but in the latter case many trivial facts may have to be gathered to make up 400~500 items. In other words, the items in the latter case would come from the population dissimilar to that of the former. In such a case, the quality of the questions in both tests would be quite different and a direct comparison of the two tests would be irrelevant. Therefore, MCQs are not a justifiable tool for the assessment of knowledge. They may still, however, be useful for self-assessment, since only the student knows which items he actually did not know.

Acknowledgments

A part of this work was supported by the Grant-in-Aid for Scientific Research, No. 62510149 from the Ministry of Education, Science and Culture in Japan and the Grant for Medical Education from Sofukai, Okayama.

REFERENCES

- 1) Pickering, G.: Against multiple choice question. *Med. Teach.* 1 : 84-86, 1979
- 2) Anderson, J.: For multiple choice questions. *Med. Teach.* 1 : 37-42, 1979
- 3) Anderson, J.: The MCQ controversy — A review. *Med. Teach.* 3 : 150-156, 1981
- 4) Eicha, L.W. : Medical-school education, 1975-1979. A student's perspective. *N. Engl. J. Med.* 303 : 727-734, 1980
- 5) Dedley, H.A.F.: Multiple-choice test: Time for a second look ? *Lancet* 2 : 195-196, 1973
- 6) Barker, D. and Maatsch, J.L.: The MCQ controversy: The issue is content, not form. *Med. Teach.* 2 : 151-152, 1980
- 7) Hubbard, J.P. and Clemens, W.V.: Multiple-choice examinations in medicine. Philadelphia, Lea and Febiger. 1961