

A Review on Writing Multiple Choice Questions

*Rahman MA,¹ Rahman H,² Islam MM³

Multiple-choice assessment is the most common type of written test item used in undergraduate, graduate and post-graduate medical education. A standard multiple-choice test item consists of two basic parts: a problem (*stem*) and a list of suggested solutions (*alternatives*). Several item-writing principles have been investigated for their effects on test psychometric indices. Most studies evaluate the effect of a single-item defect, such as negative stems and none of the above option. Effective MCQ construction requires knowledge of established MCQ writing principles. It is important for test developers to be skilled in effective MCQ writing to ensure for valid assessments of a learner's knowledge.

[Dinajpur Med Col J 2014 Jul; 7 (2):132-135]

Key words: Multiple choice question

Introduction

Multiple choice is a form of assessment in which respondents are asked to select the best possible answer (or answers) out of the choices from a list.¹ It is the most common type of written test item used in undergraduate, graduate and post-graduate medical education.² It has been demonstrated that Multiple-choice assessment is a useful device for student assessment in core university subjects that typically have large student numbers.^{3,4}

Appropriately constructed MCQs result in objective testing that can measure knowledge, comprehension, application and analysis.⁵

Structure

A standard multiple-choice test item consists of two basic parts: a problem (*stem*) and a list of suggested solutions (*alternatives*). The stem may be in the form of either a question

or an incomplete statement, and the list of alternatives contains one correct or best alternative (*answer*) and a number of incorrect or inferior alternatives (*distractors*).⁶ The "item" is the entire unit and consists of a stem and several options. The "stem" is the question, statement or lead-in to the question. The possible answers are called "alternatives", "options", or "choices." The correct option is called the "keyed response." The incorrect options are called "foils" or "distractors."

The stem is usually written first and is best written as a complete sentence or question. Direct questions (e.g., which of the following is an imaging feature of benign pulmonary nodules) are clearer than sentence completions (e.g., Benign pulmonary nodules...). Research has shown that the use of incomplete stems lowers the students' correct response rate by 10% to 15%.⁷

1. * Dr Md Abedur Rahman, Associate Professor and Head, Department of Physiology, Kushtia Medical College, Kushtia. Abed.dr.dmc@gmail.com
2. Dr Homayera Rahman, Associate Professor, Department of Physiology, Green Life Medical College, Dhaka.
3. Dr Md Montasir Islam, Associate Professor and Head, Department of Physiology, Central Medical College, Comilla

*For Correspondence

Construction

A well constructed MCQ will be unambiguous, clearly set to a defined standard and easy to mark (usually automatically), but more often than not tests little more than recall of fact.⁸ Haladyna and Downing⁹ classified 43 multiple-choice (MC) item-writing rules into three major categories: (1) general item writing, (2) stem construction, and (3) option development. The seven rules originally selected for this synthesis are those most frequently empirically studied;¹⁰ the wording of the rules is theirs. The first rule is a general item-writing procedural consideration: *avoid the complex multiple-choice format*. Two are stem construction considerations: *state the stem in question form*, and *word the stem positively*. Four are general option-development considerations: *use as many functional distractors as possible*; *avoid, or use sparingly, the phrase "all of the above;"* *avoid, or use sparingly, the phrase "none of the above;"* and *keep the length of options fairly consistent*.

Several item-writing principles have been investigated for their effects on test psychometric indices.¹¹ Most studies evaluate the effect of a single-item defect, such as negative stems¹² and none of the above option.¹³

In items of the *negative* variety, the student is directed to identify either the alternative that is an incorrect answer, or the alternative that is the *worst* answer. Any of the other multiple-choice varieties can be converted into this negative format.

Example of negative variety: Which of the following is *not* a local hormone of GIT?

- a. Secretin
- b. Gastrin
- c. CCK
- d. Vasopressin

For most educational objectives, a student's achievement is more effectively measured by having him or her identify a correct answer rather than an incorrect answer. Just because the student knows an incorrect answer does not necessarily imply that he or she knows the correct answer.

For this reason, *items of the negative variety are not recommended for general use*.⁶

All of the above and none of the above are often added as answer options to multiple-choice items. This technique requires the student to read all of the options and might increase the difficulty of the items, but too often the use of these phrases is inappropriate. None of the above should be restricted to items of factual knowledge with absolute standards of correctness. It is inappropriate for questions where students are asked to select "the best" answer. All of the above is uncomfortable in that many students will choose it if they can identify at least one of the other options as correct and therefore assume all of the choices are correct - thereby obtaining a correct answer based on partial knowledge of the concept/content.¹⁴

Several studies have indicated that including irrelevant material in the item stem decreases both the reliability and the validity of the resulting test scores¹⁰

Numerous studies have indicated that items are easier when the answer is noticeably longer than the distractors when all of the alternatives are similar in length.¹⁰ Research on the use of "all of the above" is not conclusive, the use of "none of the above" has been found in several studies to decrease item discrimination and test score reliability.¹⁰

Avoid including keywords in the alternatives. When a word or phrase in the stem is also found in one of the alternatives, it tips the student off that the alternative is probably the

answer. Several studies have reported that items are easier when a keyword in the stem is also included in the answer.¹⁰

Use as many functional distractors as are feasible.

Functional distractors are those chosen by students that *have not* achieved the objective and are ignored by students that *have* achieved the objective. In other words, they have positive discrimination.

Numerous studies have reported that there is little difference in difficulty, discrimination, and test score reliability among items containing two, three, and four distractors.¹⁰

Avoid using unnecessarily difficult vocabulary.

If the vocabulary is somewhat difficult, the item will likely measure reading ability in addition to the achievement of the objective for which the item was written. As a result, poor readers who have achieved the objective may receive scores indicating that they have not.

Use difficult and technical vocabulary only when essential for measuring the objective.

Although very little research has been done on this guideline, one study has reported that simplifying the vocabulary makes the items about 10% easier.¹⁵

State the stem in positive form (in general).

Negatively-worded items are those in which the student is instructed to identify the exception, the incorrect answer, or the least correct answer. Such items are frequently used, because they are relatively easy to construct. The teacher writing the item need only come up with one distractor, rather than the two to four required for a positively-worded item.

Positive items, however, are more appropriate to use for measuring the attainment of most educational objectives. For information on appropriate uses of negative items, see the section in this booklet entitled, "Varieties of Multiple-Choice Items"⁶

Keep the alternatives mutually exclusive.

Alternatives that overlap create undesirable situations. Some of the overlapping alternatives may be easily identified as distractors. On the other hand, if the overlap includes the intended answer, there may be more than one alternative that can be successfully defended as being the answer.⁶

Keep the alternatives homogeneous in content.

If the alternatives consist of a potpourri of statements related to the stem but unrelated to each other, the student's task becomes unnecessarily confusing. Alternatives that are parallel in content help the item present a clear-cut problem more capable of measuring the attainment of a specific objective.⁶

Conclusion

Measurement of a learner's knowledge is an important step in the educational evaluation process. Effective MCQ construction requires knowledge of established MCQ writing principles. It is important for test developers to be skilled in effective MCQ writing to ensure for valid assessments of a learner's knowledge.

References

1. Phelps, Richard. "Are US Students the Most Heavily Tested on Earth?", *Educational Measurement: Issues and Practice* 1996; 15 (3): 19–27
2. Farley JK. The multiple choice test: writing the questions. *Nurse Educator* 1989; 14:10-12, 39
3. Geiger, M. . On the *benefits of changing multiple-choice answers: Student*

- perception and performance. Education* 1996; 117(1): 108-116.
4. Milia, L D. Benefiting from multiple-choice exams: the positive impact of answer switching. *Educational Psychology* 2007; 27(5): 607 — 615
 5. Kemp JE, Morrison GR, Ross SM. *Developing evaluation instruments. In: Designing Effective Instruction. New York, NY: MacMillan College Publishing Company 1994;180-213*
 6. Burton, Steven J., Sudweeks, Richard R., Merrill, Paul F., Wood, Bud. (1991) How to Prepare Better Multiple-Choice Test Items: Guidelines for University Faculty. Brigham Young University Testing Services and The Department of Instructional Science. available at: <http://testing.byu.edu/info/handbooks/betteritems.pdf>, accessed on 26/12/2013
 7. Kent TH, Jones JJ, Schmeiser CB. Some rules and guidelines for writing multiple choice test items. Iowa City: University of Iowa College of Medicine and American College Testing Program 1974.
 8. Epstein RM: Assessment in Medical Education. *N Engl J Med* 2007; 356:387-96.
 9. Haladyna, T. M., & Downing, S. M. A taxonomy of multiple-choice item writing rules. *Applied Measurement in Education* 1989; 2(1): 37-50.
 10. Haladyna, T. M., & Downing, S. M.. Validity of a taxonomy of multiple-choice item writing rules. *Applied Measurement in Education* 1989; 2(1): 51-78.
 11. Haladyna TM, Downing, SM, Rodriguez MC. A review of multiple-choice item-writing guidelines. *Applied Meas Educ* 2002; 15:309-333
 12. Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew H. The quality of in-house medical school examinations. *Acad Med* 2002; 77:156-161
 13. Downing SM, Baranowski RA, Grosso LJ, Norcini JJ. Item type and cognitive ability measured: the validity evidence for multiple true-false items in medical specialty certification. *Appl Meas Educ* 1995; 8:189-199
 14. Gronlund, N. E., & Linn, R L. *Measurement and evaluation in teaching* (6th ed.). New York: Macmillan Publishing Company. 1990
 15. Cassels, J. R. T., & Johnstone, A. H. (1984). The effect of language on student performance on multiple-choice tests in chemistry. *Journal of Chemical Education* 1984; 61: 613-615.