UNSW
THE UNIVERSITY OF NEW SOUTH WALES

EDUCATIONAL ASSESSMENT
AUSTRALIA

MULTIPLE CHOICE

A GUIDE TO GOOD PRACTICE

NICK CONNOLLY

QAMTC 2008

Content

Multiple Choice: A guide to good practice

Multiple choice questions have a very mixed reputation. For many years now they have been popular with test designers but that rise in popularity in one group has often been met with scepticism. This paper will look at multiple choice items in the context of mathematics and will examine their strengths and weaknesses as a method of assessment.

What are multiple choice items?

The standard multiple choice item is the most common form of a style of test question called "selected response". Essentially any question where the test taker picks from a set of pre-chosen responses is a selected response item. Typically such items are multiple choice but other variants exist such true/false items or items where the test taker matches related pairs of objects.

Example 1: True false
Select true or false for each of these statements:
1. 27 is an even number. T/F
2. 13 is a prime number. T/F
3. 8 is a cube number. T/F]

Example 2: Matching
Match the measure to the unit

metres	mass
litres	time
seconds	length
kilograms	capacity

Selected response contrasts with constructed response (or free response) where the test taker creates their own response.

Example 3: CR and SR format
Constructed response: 23 - 17 = ________
Selected response: 23 - 17 = ?. (A) 6 (B) 14 (C) 16

Although the distinction seems clear there are grey areas - a question were the student is expected to give a day of the week as answer for example or a month of the year or to name one of a set of people.

Example 4:SR/CR
These are the heights of four friends: Jim 1.67 m, Ian 1.71 m, Sue 1.73 m, Lee 1.63 m.
Which friend is tallest? ____________________

Example 5:SR/CR
Which month starts 92 days after the start of March? _________________

Although even these examples do give some chance for a test taker to respond in an unexpected way (for example "Fred" or "Winter" or more revealingly "I don't know")

Structure

A test item in general is structured in three parts.
Stimulus: a text containing relevant information on the test question.
Stem: a specific question or instruction
Response area: where (or how) the student answers the question
For example the question on student heights is structured in this way:

stimulus: These are the heights of four friends: Jim 1.67 m, Ian 1.71 m, Sue 1.73 m, Lee 1.63 m.
stem: Which friend is tallest?
response area: ____________________

Although this structure is neat the distinction between each part isn't necessarily always clear nor does it need to be.

In multiple choice items the response area is much richer than in a typical constructed response item. The choices are referred to as 'options', the correct choice is called the 'key' and the incorrect options are called 'distractors' (or sometimes 'foils').

Example 6.

Sophia has two bags. In each bag there are 4 white chess pawns and 4 black chess pawns.

Without looking, Sophia takes a pawn from each bag.

What is the probability that at least one of the pawns will be black?

(A) $\frac{3}{4}$ (B) $\frac{1}{2}$

(C) $\frac{1}{4}$ (D) $\frac{2}{16}$

Stimulus and stem combined: 23 - 17 = ?

Options: Key: (A) 6 Distractors: (B) 14 (C) 16

Problems with Multiple Choice

The central problem, in my opinion, about MC items is that they are opaque. This manifests itself in several ways.

1. Guessing. You can never know whether a correct answer was a correct response or a random guess. With three option multiple choice a random choice will be correct one in three times for example. Even a 5 option multiple-choice item can have a guess rate of 20%.
2. Unclear method. Typically only one of a limited set of options is recorded. Even if the student didn't make a wholly random guess you still can't be sure they chose the correct option for the right reasons.
3. Working back from answers. Options provide additional information which may influence the way in which the student answers the problem.
4. Working is not rewarded. Largely correct working with a minor error can lead to a wrong option choice. While clear working is an important maths skill multiple-choice items do not encourage that skill.
5. Not appropriate for larger, multistage problems. Following on from point 4, more complex mathematical tasks may have many stages and students need to show a wide range of skills. A multiple-choice item essentially only rewards completion of a problem. This is a net loss of information about a student's abilities.
6. Not appropriate from problems with an open solution. Problems where students define their own problem and find optimal solutions according to their own criteria (common features of rich-tasks, projects, investigations) can't be boiled down to a narrow set of responses by their very nature.

Those six points are not the only criticisms levelled at multiple-choice. Other criticisms that have been levelled include the following.

1. Myth: Multiple-choice only test lower order thinking skills. Outside of mathematics it is true that MC format is often used for items that test basic recall of facts. I do not believe this holds true form mathematics. Mathematics problems can be extraordinarily rich in terms of thinking and yet still boil down to one distinct correct answer an a small number of plausible but incorrect responses.
2. Myth: Multiple-choice items favour boys. What evidence there is for this outside of mathematics is mixed. In subjects like History where differences have been found MC items have typically been testing different content or skills than essay questions making it hard to distinguish where the gender issue is coming from: the content or the format. The evidence in mathematics simply isn't there. Comparison of stem equivalent items in mathematics has shown little or no difference in what is being tested other than in general difficulty.
3. Myth: Multiple-choice items are easy. Multiple-choice items are generally easier than a stem equivalent constructed response item. However, 'easier' is not the same as 'easy'. The same range of ability and difficulty can generally be tested.

Objectives of good practice

Bearing all the issues in mind what objectives should inform good practice?

1. Follow general good practice in tests
2. Reduce guessing
3. Improve diagnostic information
4. Include tasks of sufficient depth

General good practice includes, clear, unambiguous items that are free from bias. A good maths test question is one that discriminates on mathematical skill and doesn't discriminate on anything else.
Language is a problem for any mathematics assessment that goes beyond straightforward arithmetic or algebra. More interesting problems require some explanation as do problems that have more substantial real world context. In all cases the language should only be as complex as it needs to be. Some good general tips:

Use short sentences even if it means using more words. Complex sentences that cover a wide range of ideas and included a variety of different facts, such as the unreliability of readability statistics when used on mathematical texts, but also include several clauses (and sub clauses), as well as range of connectives and sophisticated punctuation (eg: colons, semi-colons and hyphens), tax not only reading skills but working-memory as many things have to be remembered to make sense of whatever the final point of the sentence was.

Use people rather than passive voice: "Jim drew a circle with a radius of 20 cm." is easier to read than "A circle of radius 20 cm was drawn." The latter style of writing either shouldn't be used or reserved for students of an age where their language skills are more than competent to deal with the artificial tone or used specifically to prepare students for an external test that still uses that style.

Avoid conditionals. This is a very difficult habit to break and sometimes a sentence starting with "If..." really is necessary. However, often an "If" is redundant.

Example 7:

7.1 A circle with centre C and of radius 20 cm was drawn and two points, A and B, were marked on the circumference of the circle to form an arc. If the area of the sector of the circle formed by the arc is 15 cm^2, what is the size of angle ABC ?

7.2 Jim drew a circle with a radius of 20 cm. He marked two points on the circumference, A and B. He also marked the centre of the circle C. The area of the sector of the circle formed by A, B and C was 15 cm^2.

What is the size of angle ABC ?

The second version says much the same thing but relevant facts are packaged into short sentences which makes the information easier to find. The "if" was irrelevant in this case - a simple guideline is to ask

yourself in response to "If the area is X" "What if is isn't?" If it doesn't matter then you don't need the "if". Assert it as a fact about a particular circle (or whatever) rather than as a condition attached to some more general (and harder to imagine) circle.
Good practice also includes using items that are appropriate for the people being tested.
One last note about language applies only to multiple-choice items: avoid giving grammatical clues to the correct answer. This particularly a problem with the sentence completion style of multiple-choice item.

Example 8:
A triangle with two equal sides is an
(A) right angled
(B) equilateral
(C) isosceles
option (A) cannot be correct grammatically

In general the sentence completion format is not ideal any way and is rarely a good choice in mathematics tests.

Reduce guessing. General good practice should in itself reduce guessing. We should be trying to ensure that items engage students. This is true of all assessment items - while guessing is a particular problem for multiple-choice (because a student can get an answer correct without engaging with the item at all) it is still a problem if a student simply guesses the answer to a constructed response or simply omits the item altogether. In either case no information is gained about the students ability. Yet an issue still remains with multiple-choice items even if we produce a very clear and engaging test question.

The art of option writing.

The signature difference between a multiple-choice item and a constructed response item is the options. Any reduction in guessing and any improvement in diagnostic information has to come from the quality of the options. One option is unique, the key, but in many ways the other options (the distractors) are of equal importance. Each option can be thought of as an error trap. A student's encounter with a maths problem is not wholly unpredictable even if it is something governed as much by chance as by reason. The distractors we choose for an item to a great extent control what that item is testing and what information we can gain from it.
The distractors should cover a range of common but non-trivial errors. To write distractors involves a little cognitive psychology as you need to consider what is a reasonable way of answering a given problem incorrectly.

Plausibility

Effective distractors should be plausible. This follows from their role in hiding the key. The key should not stand out or be obviously the correct answer without regard to the skill tested. With mathematics items keys that can be deduced logically are not as problematic as keys which can be discovered because of grammatical or linguistic clues.
Some mathematics items (particularly ones with numerical answers) often appear to have distractors based on providing three numbers similar in size to the key. This is sometimes characterised as one-below, one-above and one-very close. This approach is inadequate for several reasons.
Firstly it assumes that students hunt for an answer that matches the conditions given in the item. This assumption does accord with some multiple-choice items particularly in other subjects or tests of recall. However, a better assumption is that students engage with a mathematical task and arrive at a response. If the response is not present then the student has effectively been given extra information about the problem – this can affect the performance of the item.
Another issue is key-balance. Distractors chosen in this way lead to keys being either option-B or option-C when the options are listed in numerical order.

Finally the quality of information gained from the test item is diminished when the distractors are chosen merely on resemblance. As a consequence whilst plausible distractors are needed, plausibility is not a sufficient quality for distractors for most items

Common Errors

A more sophisticated method of distractor choice is to consider the likely errors students will make. Not only does this provide meaningful distractors, it is also a useful way to consider the quality of the item in general.

Identifying the common errors involves a degree of guess work but can be made less subjective by examining the performance of similar items in previous tests. Knowledge of the skills and knowledge of the test population is also helpful.

One approach to identifying common errors is to look at the steps needed to find the key. For example the addition 27 + 26 requires the student to add several digits together whilst being aware of place value. The most common error would be to treat 7 + 6 as 3 rather than 13. If students were asked this problem as a free-response then the most common incorrect answer would be 43.

As we need three plausible distractors the task in the item has to have enough complexity to generate three errors. The addition 27 + 26 would be inadequate for most audiences. More complex problems may generate so many plausible errors that selecting the best three can be difficult.

This approach can lead to better item writing in general. Consider the flaws in this item:

Example 9

$2^2 = ?$

(A) 4 (thinks $a^2 = a + 2$)

(B) 4 (thinks $a^2 = a \times 2$)

(C) 4 (key)

Distractor Reasoning

Distractor reasoning should mirror the reasoning behind the key. It should indicate the faulty method that a student might follow to arrive at the distractor. Providing distractor reasoning forces the item writer to consider how the student may approach the question. This has a double benefit of sometimes identifying a flaw in the item – for example when a faulty method leads to the key.

The reasoning given should be sufficient to give a reviewer a clear idea of why the distractor was chosen. A short calculation (with the error indicated) is usually sufficient.

Note that when an item is edited for inclusion in a test it is sometimes necessary to change some of the parameters in the question. With clear reasoning given for all the options (the key and the distractors) it is much easier for an editor to adapt the item.

Common Distractors

For complex problems a common choice for a distractor is based on an incomplete method. That is the student has chosen a response which matches with a correct calculation at some stage of the problem rather than the final answer.

Example 10

120 students went on a trip. Two thirds travelled by bus. Each bus could carry 23 students.

How many buses were needed?

(A) 3

(B) 4

(C) 5

(D) 6

A distractor should not be based on reasoning more sophisticated than the key. Whilst the error might be a common one such a distractor will be more likely to produce wrong answers from students with better ability in mathematics. For example a possible error when calculating the hypotenuse of a right-angled triangle is to mistakenly use the method for calculating a shorter side when given the hypotenuse. Such an option could only serve to make the item perform less well as some less able students will have enough skill to find the key but insufficient skill to ever make the error used in the distractor.

Option driven items: An item that can't be answered without its options is called 'option driven'. Typically the item gives a set of conditions which apply to many objects in theory and ask which of the four options the condition applies to. In this case the distractor reasoning should indicate why the student would mistakenly think the conditions were true for that option.

Negative items: Some items ask about a condition which does NOT hold. In this case the three distractors are, in effect, true answers which do match the given condition. Such items should be used only rarely but when they are used the distractor reasoning should indicate why a student might plausibly regard the options as being FALSE.

Haladyna's 30 point checklist

The most substantial study of multiple-choice writing rules has been conducted by Thomas Haladyna of Arizona State University (Developing and Validating Multiple-Choice Test Items). From that research Haladyna produced a 30 point list of guidelines. Not all of the points are of equal weight and some have only a weak basis in research. Some are less applicable to mathematics than others.

Haladyna's "Guidelines for Writing MC Items"

- Content Concerns

1. Base each item on specific content and a type of mental behaviour.
2. Keep the specific content of items independent from one another.
3. Avoid overly specific and overly general knowledge.
4. Focus each item on a single behaviour instead of a chain of behaviours.
5. Avoid opinion-based items.
6. Avoid trick items

- Formatting concerns

7. Avoid true-false and complex MC formats.
8. Format the item vertically instead of horizontally.

- Style Concerns

9. Edit and proof items.
10. Keep vocabulary simple for the group of students being tested.
11. Use correct grammar, correct punctuation, capitalization, and spelling.
12. Minimize the amount of reading in each item.

- Writing the Stem

13. Use either a question stem or a partial sentence
14. Ensure that the directions in the stem are very clear.
15. Include the central idea in the stem instead of the choices.
16. Avoid window dressing (excessive verbiage).
17. Word the stem positively, avoid negatives such as NOT or EXCEPT.

- Writing the Choices

18. Use as many choices as possible, but three seems to be a natural limit.
19. Make sure that only one of these choices is the right answer.
20. Vary the location of the right answer according to the number of choices.
21. Place choices in logical or numerical order.
22. Keep choices independent; choices should not be overlapping.
23. Keep choices homogeneous in content.
24. Keep the length of choices about equal.
25. Avoid using non-of-the-above, all-of-the-above, or I don't know.
26. Phrase choices positively; avoid negatives such as NOT.
27. Avoid giving clues to the right answer, such as
    a. Specific determiners including always, never, completely and absolutely.
    b. Clang associations, choices identical to or resembling words in the stem.
    c. Grammatical inconsistencies that cue the test-take to the correct choice.
    d. Conspicuous correct choice.
    e. Pairs or triplets of options that clue the test-taker to the correct choice.
    f. Blatantly absurd, ridiculous options.
28. Make all distractors plausible.
29. Use typical errors of students to write your distractors.
30. Avoid humorous choices.

