# Multiple choice item construction: Avoiding constructions that reduce validity and reliability

Cynthia Brame, Assistant Director
Vanderbilt Center for Teaching

## Anatomy of a Multiple Choice Item

Test questions consist of a stem and alternatives, one of which is the answer and the remainder of which are distractors.

A 54-year-old woman, GOPO, with a BMI of 20, smokes and works as a convenience store clerk. She is seeing you because she has been having urine leakage. Which of the following in her history is a known risk factor for urinary incontinence? — **Stem**

**Distractors**
A. BMI
B. occupation type
C. parity

**Answer**
D. smoking status

**Alternatives**

Test items can be described in terms of validity, or *the degree to which they measure the learning outcomes they purport to measure*, and reliability, or the degree to which they *consistently* measure a learning outcome. To increase validity and reliability, test writers should avoid constructions that help the test-wise and constructions that test skills not central to the stated learning outcomes.

## Guidelines for writing the stem

**The stem should:**

- be meaningful by itself and should present a definite problem.
- contain only relevant material.
- be negatively stated *only* when significant learning outcomes require it.
- be a question or a partial sentence.
  - A question stem is preferable.
  - Stems with beginning and interior blanks should be avoided.

## Guidelines for writing alternatives

**Alternatives should:**

- be plausible.
- be stated clearly and concisely.
- be mutually exclusive.
- be homogenous in content.
- not include "all of the above" and "none of the above"
- be presented in a logical order (e.g., alphabetical, numerical) to avoid a bias toward certain positions.
- be free from clues about which response is correct. Specifically, the alternatives should all
  - have grammar consistent with the stem.
  - be parallel in form.
  - be similar in length.
  - use similar language.

## References

**Steven J. Burton, Richard R. Sudweeks, Paul F. Merrill, and Bud Wood.** How to Prepare Better Multiple Choice Test Items: Guidelines for University Faculty, 1991.

**Derek Cheung and Robert Bucat.** How can we construct good multiple-choice items? Presented at the Science and Technology Education Conference, Hong Kong, June 20-21, 2002.

**Thomas M. Haladyna.** Developing and validating multiple-choice test items, 2nd edition. Lawrence Erlbaum Associates, 1999.

**Thomas M. Haladyna and S. M. Downing.** Validity of a taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2(1), 51-78, 1989.

**Susan Morrison and Kathleen Walsh Free.** Writing multiple-choice test items that promote and measure critical thinking. *Journal of Nursing Education* 40: 17-24, 2001.

**National Board of Medical Examiners.** Constructing written test questions for the basic and clinical sciences. *Available at www.nbme.org*

## Other Guidelines

- As long as all alternatives are plausible, the number of alternatives can vary among items. There is little difference in difficulty, discrimination, and test score reliability among items containing two, three, and four distractors.
- Avoid complex multiple-choice problems (i.e., alternatives such as 1 and 2; 2 and 3; 1 and 3; 1, 2, and 3).
- Keep the specific content of items independent of one another.

## Additional information

- The National Board of Medical Examiners provides an excellent tutorial on writing multiple choice items. They recommend using two questions when reviewing items: "Is the item front-loaded? Can you cover the options?" These questions encourage the question writer to place key information in the stem and to construct items that an informed test-taker can answer without choices.
- The guidelines presented above help test-writers avoid constructions that tip off the test-wise or that target skills that are not central to the learning outcomes.
    - Test-wise examinees are alert to cues that indicate the correct answer. These cues may take the form of grammatical clues, "clanging" (i.e., the use of different forms of the same word in the stem and the correct answer), convergence (i.e., the use of elements of the correct answer in multiple alternatives), or logical cues (e.g., length of the answer or the use of "always" or "never" in incorrect answers). Being attentive to these cues and following the guidelines above can help in the construction of items that are more valid evaluations of the desired learning outcomes.
    - Test items that contain irrelevant material, wordy alternatives, or negative constructions have reduced validity and reliability because they test, in part, examinees' reading ability and ability to hold information in short-term memory.
    - The use of complex multiple choice and "all of the above" and "none of the above" alternatives reduces item reliability, in part because it allows examinees to use partial knowledge to arrive at a correct answer.