How to Design Effective Multiple-Choice Tests that Assess Student Learning

UF UNIVERSITY of FLORIDA
IFAS
College of Agricultural and Life Sciences

Center for Online Learning and Technology

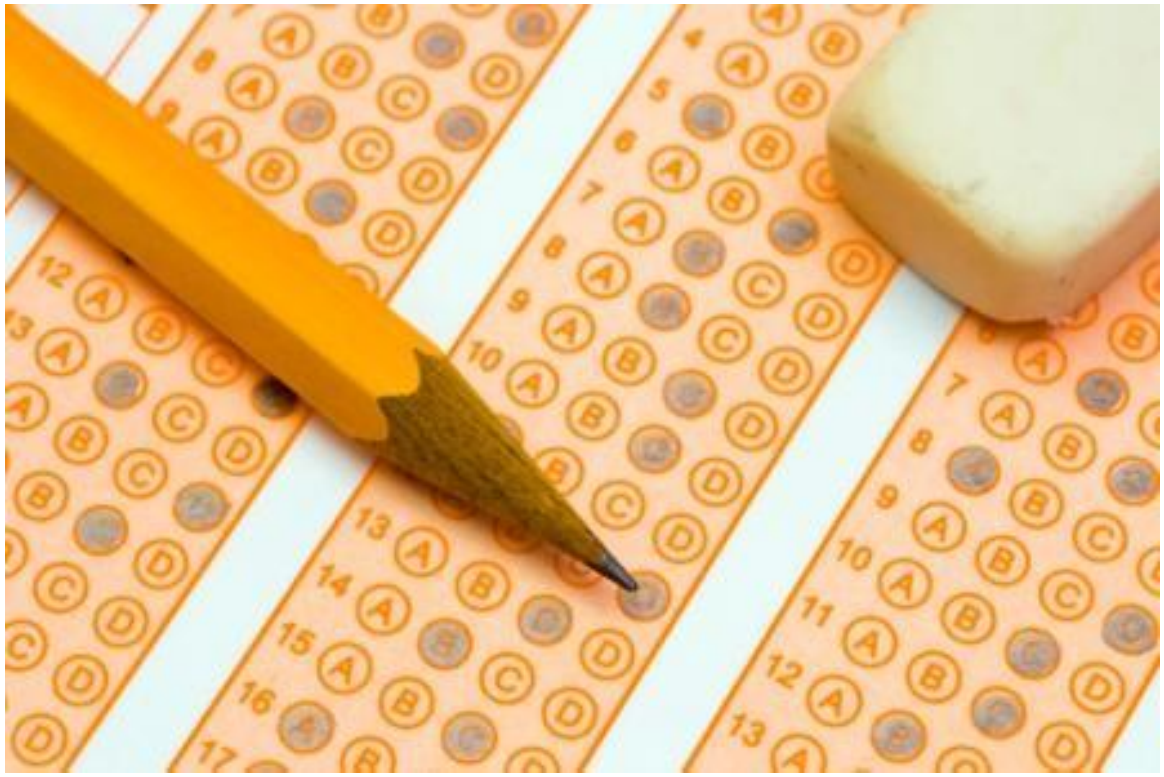
Designing Effective Multiple-Choice Tests

Objectives

By the end of this session participants will be able to:

- describe strengths and limitations of multiple-choice tests.
- evaluate appropriate uses of multiple-choice tests.
- explain guidelines for constructing multiple-choice items.
- create questions to address different levels of Bloom's Taxonomy.
- review examples of effective and ineffective multiple choice tests.
- write multiple choice questions at different cognitive levels.

About Multiple-Choice Tests



About Multiple-Choice Tests

Students select the correct answer from alternative responses. Each item has:

- item stem
- correct option
- several distractor options

Format:

- complete question – one definite answer
- incomplete question – best answer

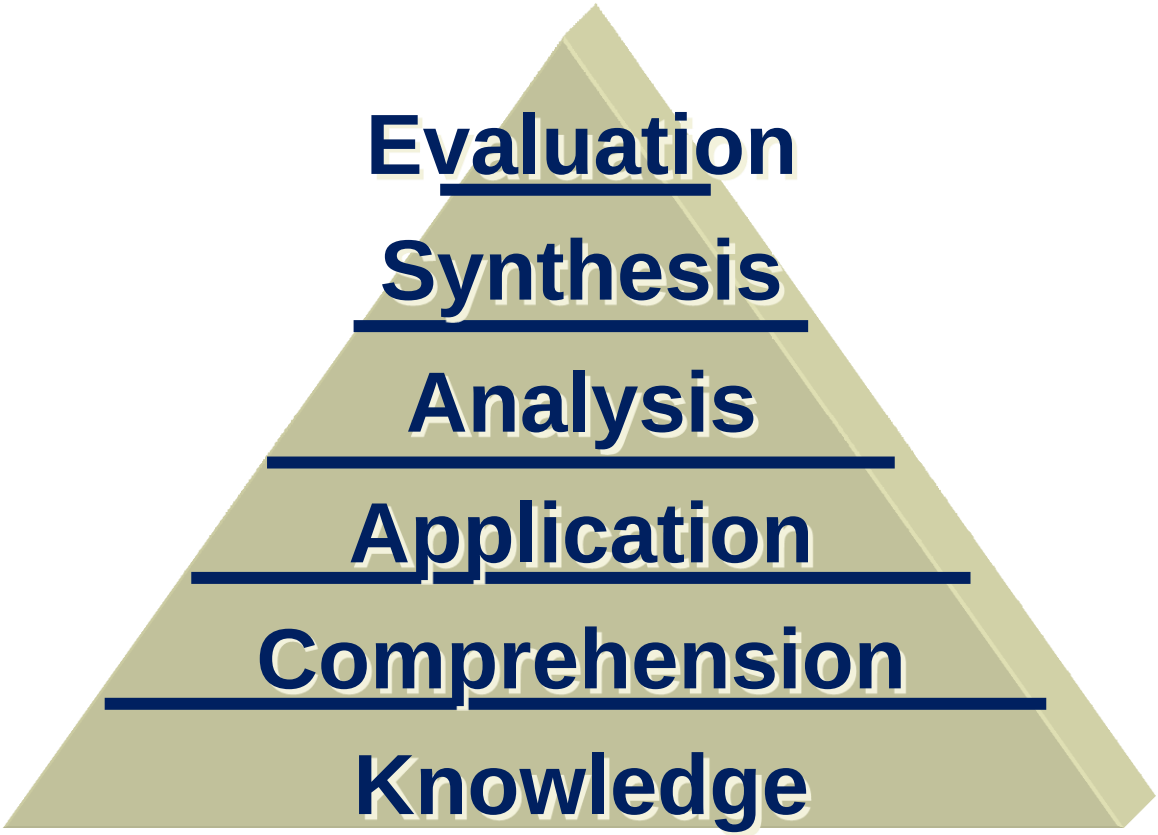
(Clegg & Cashin, 1986)

Multiple-choice Test Construction

"… the greater your experience in their construction, the longer it takes per [multiple-choice] item to construct a reasonably fair, accurate, and inclusive question."

- Wilbert J. McKeachie

Bloom's Cognitive Domain



A Resource for Question Verbs:
http://tep.uoregon.edu/resources/assessment/multiplechoicequestions/blooms.html

Advantages

Multiple-choice items can provide:

- versatility in measuring all levels of cognitive ability,
- highly reliable test scores,
- scoring efficiency and accuracy,
- objective measurement of achievement or ability,
- a wide sampling of content or objectives,
- a reduced guessing factor compared with true-false items, and;
- different response alternatives which can provide diagnostic feedback.

(Ory & Ryan, 1993)

Limitations

Multiple-choice items:

- are difficult and time-consuming to construct,
- lead an instructor to favor simple recall of facts,
- place a high degree of dependence on the student's reading ability and instructor's writing ability, and
- are particularly subject to clueing. (Students can often deduce the correct response by elimination.)

(Ory & Ryan, 1993)

When to Use

- To assess breadth of learning
- To test a variety of levels of learning
- When you have a large number of individuals taking the test
- When you have time to construct the test items
- When time is limited for scoring
- When it is not important to determine how well individuals can formulate their own answer
- When you want to prepare individuals for future assessments that use a similar format

(Clegg & Cashin, 1986)

Planning a Test

General Tips for Writing Tests

- Compose test items over time.
- Test what you really want individuals to learn.
- Check borrowed items carefully.
- Create a test bank.
- Start easy to build confidence.
- Get feedback on items.

(Nilson, 2010)

Planning a Test

- Use a test matrix or blueprint.
- Identify major ideas and skills rather than specific details.
- Use Bloom's cognitive taxonomy or something appropriate for your context.

(Nilson, 2010)

Test Matrix

Content Area to be Tested	Level of Cognition Required				Number of questions	% of test devoted to content area
	Remembering or Understanding	Applying	Analyzing or Evaluating	Creating		
Number of questions						
% of test devoted to each cognitive level						

Additional Techniques for Writing Multiple-Choice Items:

http://tep.uoregon.edu/resources/assessment/multiplechoicequestions/sometechniques.html

Objectives at Different Levels

Level: Knowledge

Objective: Identifies the meaning of a term

In the area of physical science, which one of the following definitions describes the term “polarization”?

A. The separation of electric charges by friction
B. The vibration of transverse waves in a single plane
C. The interference of sound waves in a closed chamber

Objectives at Different Levels

Level*:* Comprehension

Objective: Interprets the meaning of an idea

The statement that “test reliability is a necessary but not sufficient condition of test validity” means that:

A. a reliable test will have a certain degree of validity
B. a valid test will have a certain degree of reliability
C. a reliable test may be completely invalid and a valid test completely unreliable

Objectives at Different Levels

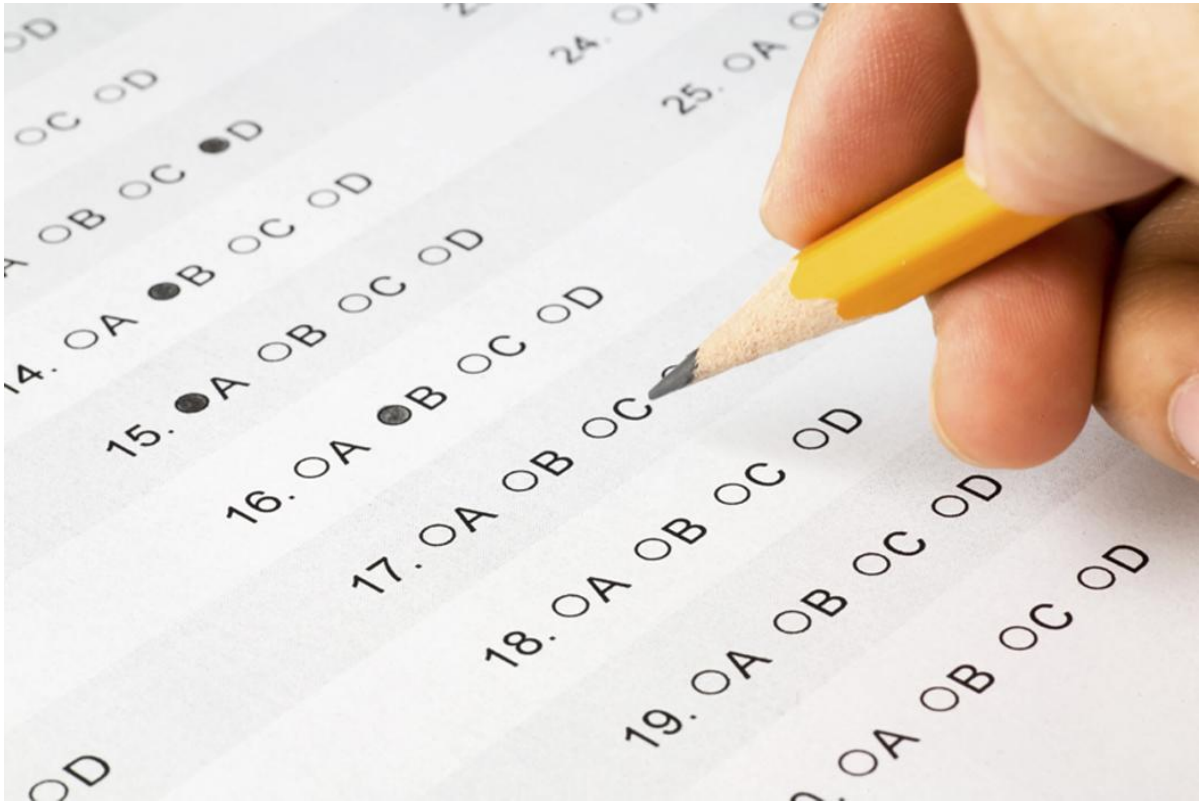
Level: Problem solving

Objective: Students must identify the correct outcome of a given circumstance

If nominal gross national product (GNP) increases at a rate of 10% per year and the GNP deflator increases at 8% per year, then real GNP:

A. Rises by 10%
B. Falls by 8%
C. Rises by 2%

Constructing Test Items



Writing Items

- Write items on significant concepts, not trivial facts.
- Write items that have a definite answer.
- Communicate clearly.
- Don’t give away the answer by including irrelevant cues in the item.
- Don’t write items that require skills or knowledge irrelevant to what you are trying to measure.
- Have items reviewed by knowledgeable persons other than the composer of the question if possible.

(Clegg & Cashin, 1986)

Components

Stem: presents the problem

Correct or keyed options: correct option

Distractor options: incorrect options

(Clegg & Cashin, 1986)

Developing an Item

1. Choose an important concept
2. Write the stem
3. Write the correct answer (key)
4. Develop distractors
   - common misconceptions
   - errors that could be made
   - plausible, yet less important information
   - similar in style, length to the key
   - every distractor should be reasonable

(Clegg & Cashin, 1986)

Issues Related to Testwiseness

- grammatical cues
- logical cues
- absolute terms
- long correct answer
- word repeats
- convergence strategy

(Clegg & Cashin, 1986)

Issues Related to Irrelevant Difficulty

- options long
- numeric data not stated consistently
- vague terms
- language not parallel
- options in no logical order
- “none of the above” is used
- stems tricky or unnecessarily complicated
- answer to an item is “hinged” to the answer of a related item

(Clegg & Cashin, 1986)

Writing Stems

- Ensure that the directions in the stem are very clear.
- Include the central idea in the stem instead of the choices.
- Avoid window dressing (excessive verbiage).
- Word the stem positively, avoid negatives such as NOT or EXCEPT. If negative words are used, use the word cautiously and always ensure that the word appears capitalized and boldface.

(Haladyna, Downing & Rodriguez, 2002)

Writing Stems

Avoid statements that fail to present a complete thought or question.

(Ory & Ryan, 1993)

Writing Stems

Avoid statements that fail to present a complete thought or question.

Eliminate excessive wording and irrelevant information.

(Ory & Ryan, 1993)

Writing Stems

Avoid statements that fail to present a complete thought or question.

Eliminate excessive wording and irrelevant information.

Include in the stem any word(s) that might otherwise be repeated in each alternative.

(Ory & Ryan, 1993)

Writing Stems

Avoid statements that fail to present a complete thought or question.

Eliminate excessive wording and irrelevant information.

Include in the stem any word(s) that might otherwise be repeated in each alternative.

Use negatively stated stems sparingly. When used, underline and/or capitalize the negative word.

(Ory & Ryan, 1993)

Writing Stems

Avoid statements that fail to present a complete thought or question.

Eliminate excessive wording and irrelevant information.

Include in the stem any word(s) that might otherwise be repeated in each alternative.

Use negatively stated stems sparingly. When used, underline and/or capitalize the negative word.

When using incomplete statements avoid beginning with the blank space.

(Ory & Ryan, 1993)

Writing Stems

Avoid statements that fail to present a complete thought or question.

Eliminate excessive wording and irrelevant information.

Include in the stem any word(s) that might otherwise be repeated in each alternative.

Use negatively stated stems sparingly. When used, underline and/or capitalize the negative word.

When using incomplete statements avoid beginning with the blank space.

Use familiar language.

(Ory & Ryan, 1993)

Writing Item Alternatives

- Develop as many effective choices as you can, but research suggests three is adequate.
- Make sure that only one of these choices is the right answer.
- Vary the location of the right answer according to the number of choices
- Place choices in logical or numerical order.
- Keep choices independent; choices should not be overlapping.

(Haladyna, Downing & Rodriguez, 2002)

Writing Item Alternatives

- Keep choices homogeneous in content and grammatical structure.
- Keep the length of the choices about equal.
- *None-of-the-above and All-of-the-above* should be used carefully.
- Make all distractors plausible.
- Use typical errors of students to write your distractors.
- Use humor if it is compatible with the teacher and the learning environment.

(Haladyna, Downing & Rodriguez, 2002)

Writing Item Alternatives

- Phrase choices positively; avoid negatives such as NOT.
- Avoid giving clues to the right answer, such as:
  - specific determiners including always, never, completely, and absolutely.
  - clang associations, choices identical to or resembling words in the stem.
  - conspicuous correct choice.
  - blatantly absurd, ridiculous options.

(Haladyna, Downing & Rodriguez, 2002)

Writing Item Alternatives

Make sure there is one correct or best response.

Which of the following does not belong with the others?

A. Wundt
B. Structuralism
C. James
D. Titchener

(Ory & Ryan, 1993)

Writing Item Alternatives

Make all alternatives plausible and equally attractive to both less-knowledgeable and skillful students.

The number of photoreceptors in the retina of each human eye is about

A. 1000.
B. 2 million.
C. 115 million.
D. 2.37 billion.

Better:

A. 5 million.
B. 35 million.
C. 65 million.
D. 115 million.

(Ory & Ryan, 1993)

Writing Item Alternatives

Minimize the use of the *all-of-the-above* and *none-of-the-above* alternatives.

Problem representation involves

A. determining which factors matter and which do not.
B. the initial state of problem solving.
C. both a and b.
D. neither a nor b.

Better:

A. determining which factors matter and which do not.
B. the initial state of problem solving.
C. reducing the problem to manageable segments.
D. all of the above.

(Ory & Ryan, 1993)

Writing Item Alternatives

Make the alternatives mutually exclusive and approximately equal in length.

Rods are found in the

A. blind spot.
B. fovea.
C. periphery or outer edge of the retina.
D. back of the eye.

Better:

A. blind spot.
B. fovea.
C. retina.
D. cornea.

(Ory & Ryan, 1993)

Writing Item Alternatives

Avoid overly wordy alternatives that become confusing and difficult to read.

Flooding differs from systematic desensitization in that

A. the former is based on classical conditioning and the latter on operant conditioning.
B. systematic desensitization requires insight and the flooding does not.
C. flooding has you start at the top of your fear hierarchy and systematic desensitization has you start at the bottom and work up gradually.
D. flooding emphasizes the use of cognitions to a much greater extent than does systematic desensitization.

Better:

Flooding differs from systematic desensitization in that flooding

A. is based on classical conditioning rather than operant conditioning.
B. doesn't require insight.
C. starts at the top of the fear hierarchy.
D. places greater emphasis on the use of cognitions.

(Ory & Ryan, 1993)

Writing Item Alternatives

Avoid irrelevant cues such as grammatical structure, well-known word associations, or connections between the stem and the correct answer.

School psychologists who examine and place children in special education settings often apply the research done by

A. biopsychologists.
B. educational psychologists.
C. clinical psychologists.
D. counseling psychologists.

Better:

School psychologists often apply the research done by

(Ory & Ryan, 1993)

Writing Item Alternatives

Avoid language that may offend or exclude a particular group of individuals.

Which of the following is a characteristic of persons with Down's syndrome?

A. Larger than normal head
B. Obesity
C. Oriental-like skin folds over the eyes
D. Above average height.

Better:

A. Larger than normal head
B. Obesity
C. Downward sloping skin folds over the eyes
D. Above average height.

(Ory & Ryan, 1993)

Critiquing Test Items

Critiquing Test Items

Twenty Thousand Leagues Under the Sea is considered to be:

A. an adventure story.
B. a science-fiction story.
C. an historical novel.
D. an autobiography.

Could be either A or B; should have one best answer.

Critiquing Test Items

When a court possesses appellate jurisdiction this means that it

A. must have a jury.
B. has the power or authority to review and decide appeals.
C. can conduct the original trial.
D. can declare laws unconstitutional.

The term “appeal” in B is too close to “appellate” in the stem.

Critiquing Test Items

Which of the following men invented the telephone?

A. Bell
B. Morse
C. Pasteur
D. Salk

C & D are not plausible distractors and the answer (A) is too obvious.

Critiquing Test Items

The indicator found by correlating students' scores on a classroom math test with their scores on a standardized math test is called a

A. validity coefficient.
B. index of reliability.
C. equivalence coefficient.
D. internal consistency coefficient.

The end of the stem is "a" which only matches answer (A).

Critiquing Test Items

In order to determine the criterion-related validity of a test, one would

A. correlate the test scores with an appropriate criterion.
B. correlate the scores from the odd and even items.
C. correlate the scores from forms a & b of the test.
D. correlate the scores from two administrations of the same test.

“Correlate the” should be included in the stem. Also both (A) and the stem have the same word, “criterion.”

Critiquing Test Items

The state that is not south of the Mason-Dixon line is

A. Mississippi.
B. Florida.
C. Kentucky.
D. Vermont.

“Not south” could trip up students and should be replaced by “north” OR the negative should be underlined or highlighted (e.g. “NOT South”). Again, answer (D) is too easy.

Critiquing Test Items

Which one of the following is the best source of heat for home use?

A. Gas
B. Electricity
C. Oil
D. Geo-thermal

“Best” is too vague. Why not use “cheaper,” “more efficient,” etc. The answer is also geographically dependent.

Critiquing Test Items

Important early theorists in the psychology of learning included

A. Ebbinghaus.
B. Thorndike.
C. Pavlov.
D. None of the above.
E. All of the above.

The stem says “theorists” so there must be more than one. (E) is the right answer. Another problem is the answer tends to be “all of the above” in this type of question. If the student can see 2 that are correct, it must be “all of the above.”

Critiquing Test Items

In a normal distribution, the mean and the median are

A. always the same point.
B. never the same point.
C. usually very close to one another.

(A) and (B) are absolutes, which are usually incorrect.
(C) is also longer.

Item Analysis

Item Analysis

- Review items for accuracy and formatting
- Have a colleague read and give feedback
- Item difficulty (percentage of students who answered each item correctly)
- Item discrimination

Summary

- Multiple-choice tests can be useful measures of learning.
- Write questions to assess the cognitive level of interest.
- Follow guidelines for writing effective multiple choice questions.
- Review student performance on items and revise exams as needed.

Questions?

UF UNIVERSITY of FLORIDA
IFAS
College of Agricultural and Life Sciences

Center for Online Learning and Technology

Resources

Clegg, V. L., & Cashin, W. E. (1986). *Improving multiple-choice tests.* Idea Paper No. 16, Center for Faculty Evaluation and Development, Kansas State University. http://www.idea.ksu.edu/papers/Idea_Paper_16.pdf.

Davis, B. G. (2009). *Tools for teaching.* (2nd ed.). San Francisco, CA: Jossey-Bass.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15(3*), 309-334.

Nilson, L. B. (2010). *Teaching at its best: A research-based resource for college instructors. (*3rd ed.) San Francisco, CA: Jossey-Bass.

Ory, J.C. & Ryan, K. E. (1993). *Tips for improving testing and grading.* Vol. 4. Newbury Park: Sage Publications.

Resources

Svinicki, M. & McKeachie, W. J. (2011). *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers.* Belmont, CA: Wadsworth, Cengage Learning.

University of Oregon, Teaching Effectiveness Program. *Writing Multiple Choice Questions that Demand Critical Thinking.* Web site: http://tep.uoregon.edu/resources/assessment/multiplechoicequestions/mc4critthink.html

University of Minnesota, Office of Measurement Services. *Writing Multiple Choice Items.* Web site: http://oms.umn.edu/fce/how_to_write/multiplechoice.php

University of Texas at Austin, Instructional Assessment Resources. *Writing Multiple Choice Items*. Web site: http://www.utexas.edu/academic/ctl/assessment/iar/students/plan/method/exams-mchoice-write.php