Creating Valid and Reliable Classroom Tests

Session II: Writing Multiple-Choice Items

Recap of Session I
Why, How, When, What of Testing
- Why should we test?
  - Discussion of purposes of testing
  - Reliability and Validity
- How should we test?
  - Different types of assessments
    - Advantages and disadvantages
  - Group vs. individual assignments
  - Take home vs. in-class assessment
  - Criterion-referenced vs. norm-referenced testing
  - Computer vs. paper-and-pencil testing
  - Test security and accommodated testing
- When should we test?
  - Frequency
- What should we test?
  - Test blueprint

Sharing of Blueprint Homework
How did the blueprint you made for your class compare with your actual test?
Was content covered in roughly the right proportions?
Was content covered by the exam that wasn’t included on the blueprint?
Was content covered by the blueprint that wasn’t included on the exam?
Other thoughts and comments? What did you learn from this experience?

Item Writing Exercise
Assemble in groups of three to complete the following:
1. Read the two items given to you and identify any problems or things you would like to change.
2. Write one MC question based on material presented so far during the Workshop.

Multiple-Choice Items
Multiple-Choice (MC) items include three components
1. Item stem
   - The part of the item that explains the basis for answering
     - Question to be answered
     - Problem to be solved
     - Incomplete statement to be completed
   The stem is followed by two or more responses (alternatives)
2. Item key
   - Correct answer
3. Item distractors
   - Incorrect choices
Types of Multiple-Choice Items

The Correct Answer Variety

- One alternative is unambiguously correct, while the others are unambiguously incorrect.
- Most straightforward and easiest to write of all types of multiple-choice items.
- Example:

  \[ 6 + 3 = \]
  a. 2
  b. 3
  c. 9
  d. 18
  Key is (c)

The Correct Answer Variety

- One alternative is unambiguously correct, while the others are unambiguously incorrect.
- Most straightforward and easiest to write of all types of multiple-choice items.
- Example:

  Reliability refers to the
  a. variation in scores.
  b. accuracy of scores.
  c. interpretation of scores.
  d. consistency of scores.
  Key is (d)

Types of Multiple-Choice Items

The Best Answer Variety

- Alternatives consist of several responses which are correct to varying degrees, or maybe completely wrong.
- Examinees are asked to select the alternative which is most nearly correct.
- Because this type involves a matter of opinion, if possible, provide the source claiming that the answer is best (e.g., text, lecture, Ben Franklin, etc.)

The best title for today's session is
a. "Strategies for Improving Classroom Testing."
b. "Introduction to Measurement."
c. "Tips for Developing Quality Multiple-Choice Tests."
d. "Assessment in the Classroom."
Key is (c)

The Best Answer Variety

- Alternatives consist of several responses which are correct to varying degrees, or maybe completely wrong.
- Examinees are asked to select the alternative which is most nearly correct.
- Because this type involves a matter of opinion, if possible, provide the source claiming that the answer is best (e.g., text, lecture, Ben Franklin, etc.)

According to Rush Limbaugh, the best choice among the 2004 Presidential candidates is
b. George Bush.
c. John McCain.
d. Wesley Clark.
e. Howard Dean.
Key is (b)

Types of Multiple-Choice Items

The Multiple-Response Variety

- Alternatives contain at least one answer which is unambiguously correct.
- Examinees are asked to select all that are correct.
- This is essentially a series of true-false questions built into a single question.

Which of the following are valid ways of assessing how well individual students have grasped the course material?
- Homework assignments
- Asking students what grade they think they deserve
- Group projects
- Essay exams
Keys are (a) and (d)

The Multiple-Response Variety

- Alternatives contain at least one answer which is unambiguously correct.
- Examinees are asked to select all that are correct.
- This is essentially a series of true-false questions built into a single question.

Which of the following is not a benefit derived from preparing a test blueprint?
- Exam questions will cover relevant course material.
- Improved reliability
- Improved validity
- Emphasis of topics on exam is appropriate.
Key is (b)

Types of Multiple-Choice Items

The Negative Variety

- Examinees are asked to select the one alternative that is incorrect.
- Occasionally useful when several good answers exist.
- It is important to be very clear in the item stem that you are interested in the one wrong answer among the alternatives.
- May be helpful to italicize or bold the negative word for emphasis.

Which of the following is not a benefit derived from preparing a test blueprint?
- Exam questions will cover relevant course material.
- Improved reliability
- Improved validity
- Emphasis of topics on exam is appropriate.
Key is (b)
Types of Multiple-Choice Items

The Substitution Variety
- A sentence is provided which contains errors in one or more places.
- Examinees are asked to identify which of the underlined parts contains the error.

The old truck battered and covered with rust, now sits behind the barn. No error:

- The old truck battered and covered with rust, now sits behind the barn. No error:
  - a. The old truck battered and covered with rust, now sits behind the barn. No error.
  - b. The old truck battered and covered with rust, now sits behind the barn. No error.
  - c. The old truck battered and covered with rust, now sits behind the barn. No error.
  - d. The old truck battered and covered with rust, now sits behind the barn. No error.
  - e. The old truck battered and covered with rust, now sits behind the barn. No error.

Key is (a)

Types of Multiple-Choice Items

The Incomplete-Alternatives Variety
- Sometimes, listing the correct answer makes the answer obvious or much easier than it would be if students were asked to produce it.

President Taylor’s first name is:
- a. James
- b. Brian
- c. Zachary
- d. Lawrence

Key is (c)

President Taylor’s first name began with what letter?
- a. A to E
- b. F to J
- c. K to O
- d. P to T
- e. U to Z

Key is (e)

Types of Multiple-Choice Items

The Combined-Response Variety
- A series of statements are made, with each statement being assigned a number.
- Examinees are asked to pick the alternative which indicates the correct relationship among the statements.

Identify correct ordering
Identify which are true

Difficult to write, but well-suited for measuring complex tasks.

For a test to be valid, it must:
- I. be internally consistent.
- II. accurately measure the construct.
- III. include items of various difficulty levels.

- a. I only
- b. II only
- c. I and II
- d. II and III
- e. I, II, and III

Key is (c)

Types of Multiple-Choice Items

The Alternate-Choice Variety
- True/False
- Right/Wrong
- Yes/No

This type of item is very hard to write well because relatively few facts are unequivocally true or false.

General Rules for Writing Test Questions
- Express the item as clearly as possible.
  - Context is very important.
  - Ambiguous, imprecise, or otherwise not understood items will not discriminate well.
  - Let the difficulty arise from the content, not the wording.

- Choose words with precise meanings.
  - Adjectives such as often, frequently, high/low, substantial, good, etc. should be avoided or clearly specified with criteria.

- Avoid complex or awkward word arrangements.
  - Use standard rules of written English.

- Include all necessary qualifications.
  - Students can’t read minds.
  - Write items so that people with different perspectives can still agree on the right answer.
General Rules for Writing Test Questions

- Avoid superfluous information
  - Students are under time constraints
  - Superfluous information detracts from the primary focus of item
  - Can cause the student to be tricked or mislead
    - Generally hurts validity
    - Can cause considerable test anxiety
- Be as accurate as possible in all parts of an item
- Make difficulty of items appropriate for group
  - Avoid using too many items that
    - all students will know
    - only the ideal students will know
  - Test should mostly include items that measure what a typical student knows
    - Test the rules, rather than the exceptions

General Rules for Writing Test Questions

- Avoid irrelevant clues to the correct response
  - Pattern among keyed response
    - Disproportionately selecting (or not selecting) an alternative as the key
    - Select keyed location first
    - Randomly assign distractors to their locations
  - Grammatical construction
    - Stem calls for plural and some alternatives are singular
    - Alternatives lead to fragments or incoherent sentences
    - Alternatives are the wrong part of speech
  - Lack of parallel structure
    - Alternatives should be uniform with respect to specificity and length
  - Repeating words in stem and key

General Rules for Writing Test Questions

- Avoid irrelevant clues to the correct response, cont’d.
  - Using specific determiners such as all, none, certainly, never, always
    - Statements including these words are almost always false.
    - Leads to easy elimination of distractors or easy True-False questions
  - Non-exclusive, synonymous, or hierarchical distractors
    - Items can’t have multiple right answers, certain choices can be eliminated as wrong.

In the United States, most people who watch television on Thursday night choose to watch
a. ER
b. Survivor
c. reality TV

General Rules for Writing Test Questions

- Interrelated items
  - The key (or a portion thereof) may appear as the stem of another item, thereby providing a clue to the right answer.

1. Which of the following is a common testing accommodation?
   a. Print exams in Braille
   b. Extended time
   c. Extra study day

8. When students request an extended time accommodation, …

General Rules for Writing Test Questions

- Avoid irrelevant clues to the correct response, cont’d.
  - Subset of alternatives that are exhaustive
    - Avoid irrelevant sources of difficulty
      - e.g., making students work with hard numbers without clean answers
    - The stem should include only one central idea
      - Avoid double-barreled questions which ask two things
        - Especially true in True-False questions where examinees may have differing opinions on two issues

George Kastanza enjoyed watching sports and writing letters  T / F
Rules for Writing Multiple-Choice Items

- Use either a direct question or an incomplete statement as stem.
  - Question format is easiest way to explicitly state the basis on which to respond.
  - Omissions in incomplete statements should occur toward end of item.
  - Avoids confusion and excess reading (and re-reading).

  **Poor:** _______ is a primary advantage of computerized testing.
  **Better:** A primary advantage of computerized testing is _______.
  **Best:** Which of the following is a primary advantage of computerized testing?

Rules for Writing Multiple-Choice Items

- Item stem should include the central problem.
  - The examinee should not have to construct the question by consulting the options.

  **The President of the United States**
  - Approves Congress’ selection of Supreme Court Judges.
  - Determines the constitutionality of laws.
  - Is elected directly by the people.
  - *must have been born in the United States.*

  This is essentially four True-False items, but only one is keyed true.

- As a rule, a good way to construct multiple-choice items is to think of giving the item as an open-ended short-answer question.
  - Stem must be sufficiently clear to have one (or very few) right answers.

Rules for Writing Multiple-Choice Items

- Include in the stem any words that otherwise must be repeated in each of the alternatives.

  One difference between criterion-referenced (CR) and norm-referenced (NR) testing is that, in CR testing,
  - the item difficulties are mostly the same.
  - the item difficulties vary widely to cover the entire achievement spectrum.
  - the item difficulties are usually set to be fairly easy.
  - the item difficulties are targeted at certain pivotal points along the scale.

  The words “the item difficulties” are contained in each of the alternatives.

Rules for Writing Multiple-Choice Items

- Avoid highly technical distractors.
  - The level of information required to reject an incorrect answer should be no higher than that required to select the correct answer.
  - Don’t let distractors detract from objective of item.

  **sad:**happy; stupid:________
  - **a.** vacuous
  - **b.** sagacious
  - **c.** jocose
  - **d.** obtuse

  Key: **b**

  Item doesn’t measure only analogical thinking anymore.
Rules for Writing Multiple-Choice Items

- Use "None of the above" with great caution.
  - Use only with correct-answer questions, never for best-answer questions.
  - Potentially useful for mathematics, spelling, grammar, etc. where correctness can be applied rigorously.
  - Use it as an obvious correct answer early in test
  - Establish that it is a viable alternative
  - Not all students seriously consider “none of the above”
  - Use only when distractors encompass most of the plausible incorrect alternatives.
  - Can be used as a key to allow instructor to avoid stating an answer which is too-obviously correct.

Use "None of the above" with great caution. Use only with correct-answer questions, never for best-answer questions. Potentially useful for mathematics, spelling, grammar, etc. where correctness can be applied rigorously. Use it as an obvious correct answer early in test. Establish that it is a viable alternative. Not all students seriously consider “none of the above”. Use only when distractors encompass most of the plausible incorrect alternatives. Can be used as a key to allow instructor to avoid stating an answer which is too-obviously correct.

Rules for Writing Multiple-Choice Items

- Arrange alternatives in a logical order
  - order of magnitude, temporal sequence, numerical, hierarchical, etc.

- Punctuate the options correctly.
  - If the stem is incomplete statement, each option should begin with a lowercase letter and end with a period.
  - If the stem is a question, alternatives should begin with a capital letter and end with a period if they are complete sentences.
  - Don’t include punctuation at the end of the stem unless it is grammatically correct
    - no colons at end of stem.

Review of Sample Items

1. The negative type of multiple-choice item in which, instead of asking students to select the one correct answer, students are asked to select the one alternative that is incorrect, is useful when
   a. attempting to determine if the students are reading the item carefully.
   b. several good answers exist
   c. attempting to determine if a student is following directions.
   d. the negative in the stem is undefined and bolded.

Problems with Item 1
1. Stem is overly wordy.
2. Alternatives (a) and (c) are very close to each other; one should probably be replaced.
3. Alternative (b) needs a period at the end.
4. Alternative (c) is singular, while the stem is plural.
5. Alternative (d) is tricky, and relies on students distinguishing between “useful” and “permissible.”
6. Ambiguous perspective: “useful” for whom?

Review of Sample Items

2. Various item formats have specific advantages and limitations. An advantage the essay format has over the multiple-choice format is
   a. the essay item can assess students’ ability to evaluate ideas.
   b. the essay item can assess students’ ability to evaluate ideas.
   c. the essay item can be reliably scored.
   d. the essay item requires students to communicate ideas.

Problems with Item 2
1. Stem is overly wordy–the lead sentence is unnecessary.
2. The words “the essay item” repeat in each alternative.
3. Alternative (a) is too vague.
4. Alternatives (b) and (d) are both correct.
Review of Sample Items

2. Which one of the following is an advantage of essay items over multiple-choice items?
   a. Assess more skills in a given amount of time.
   b. Test students’ memory of key facts.
   c. Facilitate reliable scoring of answers.
   d. Evaluate students’ ability to communicate ideas.

Key is (d)

Item Writing Exercise

Re-assemble in same groups as earlier
1. Review the item that you wrote and make any final revisions
2. Sharing of items as a group

Preview of Session III

• Homework
  1. Write a constructed-response question (e.g., essay or short answer) based on material covered in Sessions I and II.
  2. Complete 15-item MC quiz and return at next class (or by campus mail to Jim Wollack, 373 Educational Sciences).

• Next class will concentrate on rules for writing and scoring constructed-response items.