Getting the Most Out of Multiple-choice Questions

David DiBattista, Ph.D.
Professor of Psychology
Brock University
3M National Teaching Fellow–2007

Presented at York University
Toronto, Ontario
November, 2007

Contact information
David DiBattista, Ph.D.
Brock University
Department of Psychology
St. Catharines, ON L2S 3A1

E-mail: David.DiBattista@brocku.ca
Phone: 905-688-5550, ext. 3467

Tips for Constructing Multiple-choice Items

- Use question format whenever possible and sentence-completion format only when necessary. In either case, use appropriate punctuation and capitalization.
- If you use completion format, always place the blank at the very end of the stem.
- The stem should present the issue under consideration *clearly* and contain as much information as possible.
- Do not include irrelevant information in the stem unless it plays a role in the assessment procedure.
- Check carefully for spelling errors, giving special attention to distracters.
- If you use sentence-completion format, check carefully for grammatical consistency of stem and alternatives.
- Whenever possible, avoid negative wording in the stem, and be sure to emphasize it when it does occur.
- All distracters should be plausible.
- Four alternatives will usually be quite adequate, but the number used is best determined by the number of *plausible* distracters you can supply.
- To generate plausible distracters:
    1. Use students' most common errors on constructed-response tests.
    2. Use distracters that are similar to the correct answer in content, length, and complexity.
    3. Use words that sound important or have associations to the stem.
    4. Use distracters that are true, but do not correctly answer the question.
- Avoid patterns in the length and location of correct answers that could provide clues that are unrelated to content.
- Balance the answer key so that the correct response appears in each position about the same number of times.
- Do not use "none of the above."
- Do not use "all of the above" unless there are only two distracters.
- Do not use structurally complex items (e.g., "Type K"), which may call heavily on test-taking skills.
- Avoid trivia. Test only important information, concepts, and abilities.
- Ignore any of the preceding suggestions when you have a good reason to do so.

Referring to "Tips for Constructing Multiple-Choice Items," find at least one way in which each of the following items could be improved. Correct answers are italicized.

1. Sodium is the most abundant of the alkali metals, and the most common sodium compound is sodium chloride. A sodium chloride solution is said to be _____ when its concentration exceeds 0.9% (w/v).
   A. hypertrophic
   B. hyperplastic
   C. hypertonic
   D. None of the above

2. The Stanley Cup
   A. Is made of pewter
   B. Was made in Englnd
   C. Has been under the control of the National Hockey League since 1946, when there were only six teams in the league
   D. Were first presented in 1924

3. Which of these men was not a polar explorer?
   A. Robert Peary
   B. Richard Byrd
   C. David Livingstone
   D. All of the above

Bloom's Taxonomy, Revised

COGNITIVE PROCESS DIMENSION

1. **REMEMBER**: Retrieving relevant knowledge from long-term memory
   - Recognizing; Recalling
2. **UNDERSTAND**: Determining the meaning of instructional messages, including oral, written, and graphic communications
   - Interpreting; Exemplifying; Classifying; Summarizing; Inferring; Comparing; Explaining
3. **APPLY**: Carrying out or using a procedure in a given situation
   - Executing; Implementing
4. **ANALYZE**: Breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose
   - Differentiating; Organizing; Attributing
5. **EVALUATE**: Making judgments based on criteria and standards
   - Checking; Critiquing
6. **CREATE**: Putting elements together to form a novel, coherent whole or make an original product
   - Generating; Planning; Producing

KNOWLEDGE DIMENSION

A. **FACTUAL KNOWLEDGE**: The basic elements that students must know to be acquainted with a discipline or solve problems in it
   - Knowledge of terminology
   - Knowledge of specific details and elements

B. **CONCEPTUAL KNOWLEDGE**: The interrelationships among the basic elements within a larger structure that enable them to function together
   - Knowledge of classifications and categories
   - Knowledge of principles and generalizations
   - Knowledge of theories, models, and structures

C. **PROCEDURAL KNOWLEDGE**: How to do something; methods of inquiry, and criteria for using skills, algorithms, techniques, and methods
   - Knowledge of subject-specific skills and algorithm
   - Knowledge of subject-specific techniques and methods
   - Knowledge of criteria for determining when to use appropriate procedures

D. **METACOGNITIVE KNOWLEDGE**: Knowledge of cognition in general as well as awareness and knowledge of one's own cognition
   - Strategic knowledge
   - Knowledge about cognitive tasks, including appropriate contextual and conditional knowledge
   - Self-knowledge

Anderson, L.W., & Krathwohl, D., (Eds.). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's Taxonomy of educational objectives.* New York: Longman, 2001.

Bloom's Taxonomy, Revised

The Knowledge Dimension

The Cognitive Process Dimension	A. Factual	B. Conceptual	C. Procedural	D. Metacognitive
1. Remember				
2. Understand				
3. Apply				
4. Analyze				
5. Evaluate				
6. Create				

Item Shells

Haladyna and Shindoll (1989) describe an item shell as a "hollow" MC item that has a syntactic structure, but no content. The test writer can insert important concepts into the item shell to construct challenging MC items. The use of item shells can make writing MC items easier and assist the writer in constructing challenging items. The following item shells are adapted from Haladyna (1997 and 1999; see *Further Reading on Multiple-Choice Testing* for further details).

Which best defines X?

Which is the meaning of X?

Which is synonymous with X?

Which is like X?

Which is characteristic (or uncharacteristic) of X?

Which is a defining characteristic of X?

Which is an example of X?

Which statement best exemplifies the principle of X?

Which is the cause of (or reason for) X?

Which is the relationship between X and Y?

A is to B as C is to which of the following?

Which is an example of the principle of X?

If X occurs, which is most likely to be the result?

Which is most commonly the cause of X?

Which distinguishes X from Y?

Which is most (or least) important (or significant, effective, etc.)?

Which is best (or worst, or highest/lowest, biggest/smallest, etc.)?

Which is most(or least) X?

Which is a difference (or similarity) between X and Y?

Which of the following principles applies to evaluating X?

Which is the most important factor contributing to X?

Which is a major shortcoming of X?

Problem presented. Which procedure (or strategy) would be used to solve this problem?

Problem presented. Which is a possible solution?

Problem presented. Why is X the most effective (or efficient) solution?

Interpretive Exercise #1

Instructions: Read the following research scenario and answer the questions that follow.

Madame Clousseau claims to be a psychic–that is, she claims to be able to predict future events with a level of accuracy better than chance. To examine her claim, Professor Jones brings her into his laboratory and tests her under carefully controlled conditions. He tosses a standard, fair coin 300 times and has Madame Clousseau predict what the outcome will be for each toss. He finds that she correctly predicts the outcome for 157 of the tosses. When he carries out the statistical test to analyze the results, Professor Jones lets alpha equal 0.05 and he uses a two-tailed test.

1. What statistical test should Professor Jones use to analyze the data?
   A. the two-sample *t*-test for independent samples
   *B. the *z*-test for binomial probability
   C. the one-sample *z*-test
   D. the one-sample *t*-test

2. What is the null hypothesis for the statistical test?
   A. $\mu_0=150$
   B. $\mu_0=157$
   C. $\mu_1-\mu_2=0$
   *D. $\pi=0.50$

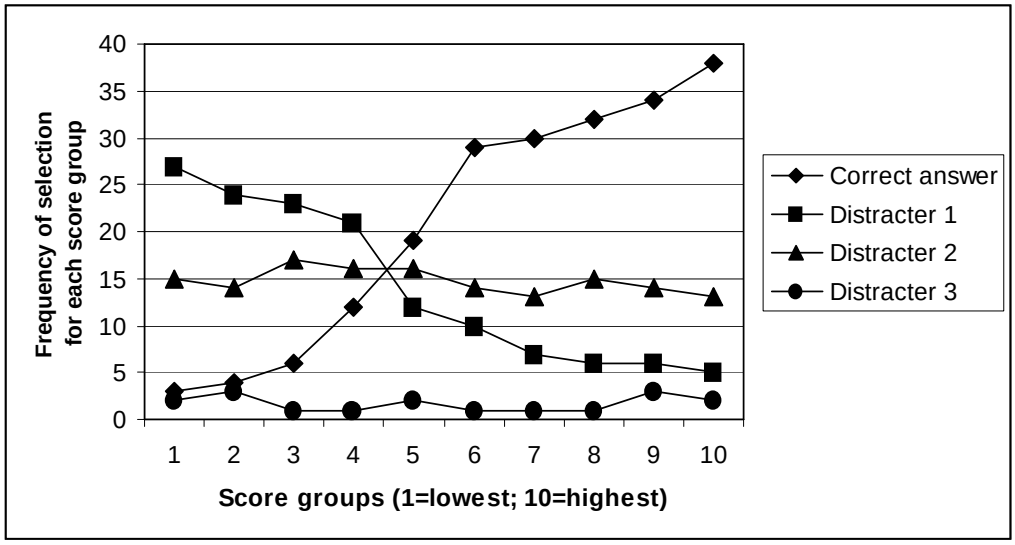
3. Suppose that the null hypothesis is actually **true**. What is the probability that Professor Jones will make a Type I error?
   *A. 0.05
   B. 0.10
   C. 0.90
   D. 0.95

4. Suppose now that the null hypothesis is actually **false**. If Professor Jones tossed the coin only 50 times rather than 300 times, what effect would this have on the power of the statistical test?
   A. The power would increase.
   *B. The power would decrease.
   C. The power would not be affected at all.
   D. This question cannot be answered using the available information.

5. Suppose once again that the null hypothesis is actually **false**. If Professor Jones set alpha at 0.10 instead of at 0.05, what effect would this have on the power of the statistical test?
   *A. The power would increase.
   B. The power would decrease.
   C. The power would not be affected at all.
   D. This question cannot be answered using the available information.

Interpretive Exercise #2

Instructions: The trace lines in the graph shown below were derived from a four-choice multiple-choice item. Answer the questions that follow.



1. Which of the following characteristics of the correct answer is particularly desirable?
   A. It is selected more frequently than any of the distracters.
   B. It is selected more frequently than all of the distracters put together.
   *C. High-scoring students are more likely than low-scoring students to choose it.
   D. Students with very little knowledge of the topic are likely to answer correctly by guessing.

2. How would Distracter 1 be characterized?
   A. plausible and a poor discriminator
   *B. plausible and a good discriminator
   C. implausible and a poor discriminator
   D. implausible and a good discriminator

3. How would Distracter 2 be characterized?
   *A. plausible and a poor discriminator
   B. plausible and a good discriminator
   C. implausible and a poor discriminator
   D. implausible and a good discriminator

4. Of the following values, which is *closest* to the total number of students that selected Distracter 2?
   A. 25
   B. 50
   C. 100
   *D. 150

Interpretive Exercise #3

Instructions: Fill in all of the values missing from the following ANOVA summary table. Each correct answer earns one point.

Source of variation	*SS*	*df*	*MS*	*F*
Between cells				
Variable A	600		300	
Variable B		1	100	
A×B				20.0
Within cells	420		10	
Total				

When you have completed the ANOVA table, answer the questions that follow. Each correct answer earns three points.

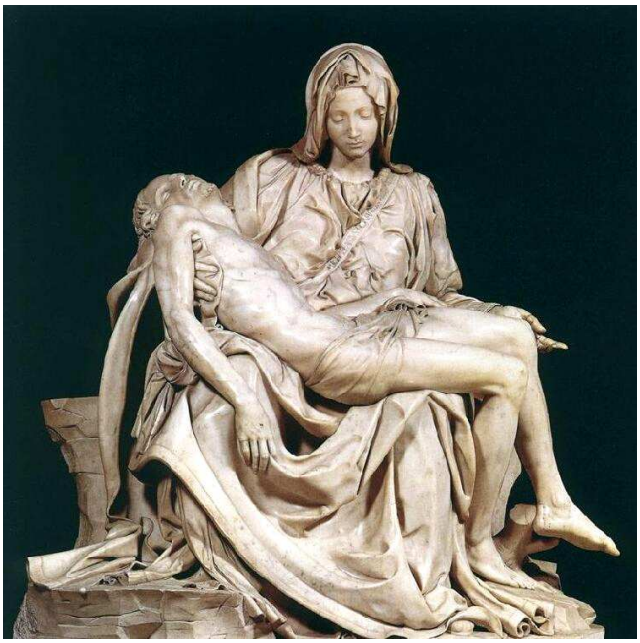
1. What type of ANOVA was used to generate this table?
   A. two-way repeated-measures ANOVA
   *B. two-way independent-groups ANOVA
   C. two-way between-within ANOVA
   D. two-way mixed-design ANOVA

2. What is the total number of scores in the data set that was used to generate this table?
   *A. 48
   B. 52
   C. 53
   D. 104

3. How many levels of Variable A are there?
   A. 1
   B. 2
   *C. 3
   D. This question cannot be answered using the available information.

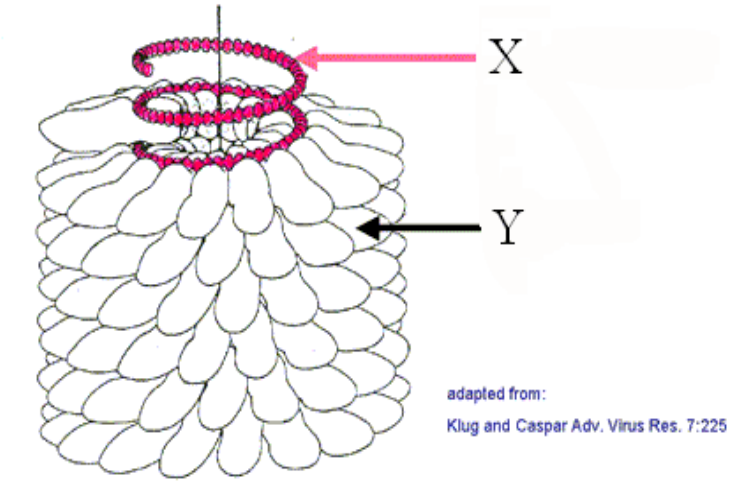
4. What is the decision rule that relates to the *F*-statistic that has the **smallest** value in this table?
   A. If $F \geq 3.22$, reject the null hypothesis.
   B. If $F \geq 3.46$, reject the null hypothesis.
   *C. If $F \geq 4.07$, reject the null hypothesis.
   D. If $F \geq 5.07$, reject the null hypothesis.

Interpretive Exercise #4



Pieta
Michelangelo Buonarotti (1475-1564)

Interpretive Exercise #5



Tobacco mosaic virus. X=nucleic acid; Y=protein.

The Cool Web

Children are dumb to say how hot the day is,
How hot the scent is of the summer rose,
How dreadful the black wastes of evening sky,
How dreadful the tall soldiers drumming by.

But we have speech, to chill the angry day,
And speech, to dull the rose's cruel scent.
We spell away the overhanging night,
We spell away the soldiers and the fright.

There's a cool web of language winds us in,
Retreat from too much joy or too much fear:
We grow sea-green at last and coldly die
In brininess and volubility.

But if we let our tongues lose self-possession,
Throwing off language and its watery clasp
Before our death, instead of when death comes,
Facing the wide glare of the children's day,
Facing the rose, the dark sky and the drums,
We shall go mad and die that way.

Robert Graves (1895-1985)

Instructions: Choose the best answer for each of the following items.

1. The poetic form of "The Cool Web" is best characterized as
   A. free verse with a concluding rhymed couplet.
   B. ballad stanzas with irregular rhymes.
   C. blank verse with unusually irregular rhythm.
   D. partially rhymed quatrains with a concluding sestet.

2. To emphasize the adult loss of childhood experience, the speaker of the poem
   A. relies on frequent breaks in the middle of the line.
   B. establishes a tone of caution, nostalgia and forgetfulness.
   C. uses imagery of drowning.
   D. alludes to the classical myth of endless return.

Continued on next page...

3. The "web of language" is ***cool*** because, according to the poem, language
   A. is the means by which heated conflict may be resolved.
   B. lessens the likelihood of achieving spiritual vision.
   C. makes our register of the world less intense.
   D. entangles us in misunderstandings.

4. When the speaker says that children "are dumb," he means that they
   A. experience life directly without the mediation of speech.
   B. relate to the world with imagination rather than intellect.
   C. lack the powers of rationality to comprehend the nuances of life.
   D. become easily overpowered by the strength of their emotions.

5. According to the speaker, which of these is a time when adults might throw off the web of language?
   A. At the moment of heightened passion.
   B. At the moment of death.
   C. At the moment of insight.
   D. At the moment of belief.

6. Which of the following, according to the speaker, would be a consequence if adults were to "[throw] off language"?
   A. Despair
   B. Absurdity
   C. Death
   D. Madness

7. Since the poem decries the limitations of language, it is paradoxical that the speaker should rely on language so effectively. Arguably, this paradox is resolved by all of the following features–**except one**. Which one does **NOT** belong?
   A. The speaker is speaking as an adult anyway, so the poem readily admits its intellectualization.
   B. The speaker uses relatively simple and direct language, so he can hardly be accused of indulging in "volubility."
   C. The tone of the poem is unpretentious, and by using "we," the speaker implicitly acknowledges his limitations.
   D. The speaker makes a special claim for "the poet," who is closer to the direct and emotional experiences of the child.

...etc.

Courtesy of Professor Brent MacLaine, Department of English, University of Prince Edward Island. Professor MacLaine uses MC items such as these in a team-learning context, with teams consisting of 4-6 students. Before coming to class, students read and study a piece of literature, such as "The Cool Web." When they come to class, students first take a MC individually and submit it for grading. They then retake the same test, but this time working as a team. Team members must work out any disagreements they may have, and they must select one single response for each item. The team then submits a single answer sheet for grading. For each student, the higher of the two marks–individual test or team test–is the one that counts. Later, the entire class engages in an extended discussion of the poem, building on the insights gained in the team-learning setting.

Tips for Interpretive Exercises

- Ensure that the introductory material is relevant to the course objectives and appropriate to students' curricular experience.
- Ensure that the introductory material is novel, and is as brief and clearly written as possible.
- Construct MC items that require analysis and interpretation of the introductory material, not simple fact-finding.
- Let the number of MC items be roughly proportional to the length of introductory material.
- Interpretive exercises can be used not only with MC items, but also with other response formats, including constructed-response formats.
- To reuse an interpretive exercise, create new introductory material, and make whatever minimal changes are necessary to the individual MC items.

Further Reading on Multiple-Choice Testing

These books focus specifically on multiple-choice testing.

Haladyna, T.M. (1999). *Developing and Validating Multiple-Choice Test Items,* 2nd edition. Mahwah, New Jersey: Lawrence Erlbaum Associates.

McDonald, M.E. (2002). Systematic assessment of learning outcomes: Developing multiple-choice exams. Sudbury, MA: Jones and Bartlett.

These books provide background on assessment techniques in general and also have helpful information specifically about multiple-choice testing.

Ebel, R.L., & Frisbie, D.A. (1991). *Essentials of Educational Measurement*, 5th edition. Englewood Cliffs, New Jersey: Prentice-Hall.

Linn, R.L. & Gronlund, N.E. (1995). *Measurement and assessment in teaching*, 7th edition. Upper Saddle River, New Jersey: Prentice-Hall.

This book focusses specifically on assessing higher-order thinking.

Haladyna, T.M. (1997). *Writing Test Items to Evaluate Higher Order Thinking.* Boston: Allyn and Bacon.

These journal articles examine the guidelines for writing multiple-choice items.

Haladyna, T.M., & Downing, S.M. (1989a). A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education, 2*, 37-50.

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

Haladyna, T.M., Downing, S.M., and Rodriguez, M.C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15,* 309-344.

A novel response technique: The Immediate Feedback Assessment Technique

This journal article presents the Immediate Feedback Assessment Technique (IFAT), a novel multiple-choice response form that uses an answer-until-correct format and gives students immediate, corrective, item-by-item feedback while they take the test. Research has shown that the IFAT promotes learning and that students strongly prefer it to other response formats, such as the more widely used Scantron form. The following article provides further information and practical tips for instructors who might like to try the IFAT:

DiBattista, D. (2005). The Immediate Feedback Assessment Technique: A learner-centered multiple-choice response form. *Canadian Journal of Higher Education, 35,* 111-131.