

# Getting the Most Out of Multiple-choice Questions

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



**Presented at Brock University**  
**February, 2014**

## The Why and The How of Effective Testing

### A primary goal of classroom testing

To measure the extent to which students have learned the facts, concepts, procedures, and skills that have been taught in the course.

### An effective test

Students who have learned more will obtain higher test scores, and students who have learned less will obtain lower scores. To be effective, a test must consist of effective items.

### What is an effective test item?

For an individual MC test item to be effective, students with higher test scores must be more likely to answer it correctly than those with lower scores.

	<u>Percent answering item correctly</u>		
	Top quartile	Bottom quartile	Top - Bottom
A good item			
A poor item			
An awful item			

## Tips for Constructing Multiple-choice Items

- When writing the stem, use question format rather than sentence-completion format.
- The stem should be written as clearly and concisely as possible, have a definite focus, and contain as much information as possible.
- Do not include irrelevant information in the stem unless it plays a role in the assessment procedure.
- Whenever possible, avoid using negative wording (*no*, *not*, *never*, etc.) when posing the question in the stem, and be sure to emphasize negative wording when it does occur.
- Use no more than four options, and ensure that all distractors are plausible.
- To generate plausible distractors:
  - ▶ Use students' most common errors and misconceptions.
  - ▶ Use distractors that are similar to the keyed option in content, length, and complexity.
  - ▶ Use words that sound important or have associations to the stem.
  - ▶ Use distractors that are true, but that do not correctly answer the question posed.
- Balance the answer key so that the keyed option appears in each position (A, B, C, and D) about the same number of times.
- Avoid providing clues that will help testwise students to find the keyed option based on item structure rather than on their knowledge of content. Examples:
  - ▶ Testwise students know that spelling and grammatical errors are far more likely to occur in distractors than in the keyed option.
  - ▶ Testwise students know that the keyed option is often substantially longer, more complex, or more detailed than the distractors.
  - ▶ Testwise students know that when the options include a single pair of opposites, the keyed option is almost always a member of the pair.
- Do not use “none of the above.”
- Do not use “all of the above.”
- Ensure that test items deal with important facts, concepts, procedures, and skills that are related to the course learning objectives.

# Bloom's Taxonomy, Revised

## The Knowledge Dimension

		Factual	Conceptual	Procedural	Metacognitive
The Cognitive Process Dimension	Remember				
	Understand				
	Apply				
	Analyze				
	Evaluate				
	Create				

**If assessment tasks are to tap higher-order cognitive processes, they must require that students cannot answer them correctly by relying on memory ALONE.**

—Anderson and Krathwohl, 2001, page 71 (emphasis supplied)

continued on next page...

## **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

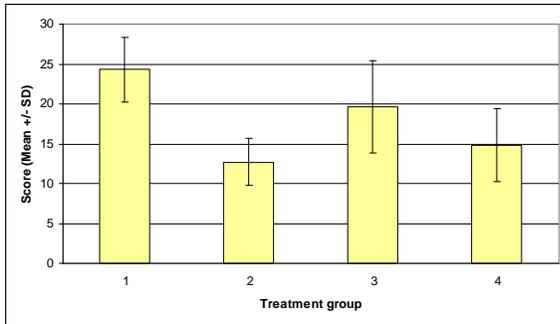
## **KNOWLEDGE DIMENSION**

- 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

# Assessing Higher-level Cognitive Processes Using Multiple-Choice Items

## **Understand: Interpreting**

In the figure shown below, which treatment group has the most variability in its scores?



- A. Group 1
- B. Group 2
- C. Group 3\*
- D. Group 4

## **Understand: Exemplifying**

Which of the following animals provides the best example of radial symmetry?

- A. a turtle
- B. a jellyfish\*
- C. a snake
- D. a snail

## **Understand: Classifying**

Right after a rat smells menthol, it is always given Drug X, which reliably induces substantial water intake. Eventually, the rat drinks water whenever it smells menthol, even when it is not injected with Drug X. In this situation, what is the role of Drug X?

- A. conditioned stimulus
- B. activational stimulus
- C. unconditioned stimulus\*
- D. discriminative stimulus

## **Understand: Classifying**

You are reading an article in which the world's major cities are ranked with respect to the quality of life for their residents. This is an example of what type of measurement scale?

- A. ordinal scale\*
- B. nominal scale
- C. ratio scale
- D. interval scale

**Understand: Summarizing**

Which of these statements best summarizes Gilligan's response to Kohlberg's theory of moral development?

- A. His theory is sound, and his research on cultural differences is particularly compelling.
- B. His theory is flawed, primarily because it ignores cultural differences.
- C. His theory is sound, and his research on gender differences is particularly compelling.
- D. His theory is flawed, primarily because it ignores gender differences.\*

**Understand: Inferring**

Vision is to Cranial Nerve 2 as hearing is to Cranial Nerve \_\_\_\_.

- A. 4
- B. 6\*
- C. 8
- D. 11

**Understand: Comparing**

Consider the measurement scales represented by (i) Social Insurance Numbers and (ii) annual incomes. How are these two scales similar to each other?

- A. Counting is meaningful with both scales.\*
- B. Subtraction is meaningful with both scales.
- C. Multiplication is meaningful with both scales.
- D. Forming ratios is meaningful with both scales.

**Understand: Explaining**

Why is the  $z$ -test for comparing means of independent samples so rarely used by behavioural scientists?

- A. The test requires equal sample sizes, and it rarely happens that sample sizes are equal.
- B. The null hypothesis that it tests is rarely of interest to behavioural scientists.
- C. It has much less power than other procedures that can be used to test the same null hypothesis.
- D. It rarely happens that the population variances are known values.\*

**Apply: Executing**

Working with an ordinal data scale, Jeff obtained the following five scores: 0, 0, 2, 5, 8. What is the value of the median for this set of scores?

- A. 0
- B. 2\*
- C. 3
- D. 5

**Apply: Implementing**

A researcher is planning an experiment in which each participant will receive both Treatment A and Treatment B. He is now thinking about the order in which the two treatments will be given to participants. Which of the following would be most advisable?

- A. Pick one order at random, either A-B or B-A, and give all participants the treatments in that order.
- B. For each participant, let the research assistant decide the order in which to give the treatments.
- C. Let participants decide for themselves the order in which they would prefer to be given the treatments.
- D. Give the treatments to half of the participants in the order A-B, and to the other half in the order B-A.\*

**Apply: Implementing**

Working with an ordinal data scale, Jeff obtained the following five scores: 0, 0, 2, 5, 8. What is the value of the most appropriate measure of central tendency for this set of scores?

- A. 0
- B. 2\*
- C. 3
- D. 5

**Analyze: Differentiating**

Keri received a grade of 70 on her history test. 200 people took the test, and scores ranged from a low of 30 to a high of 90. The class mean was 60, and the variance was 100. Which numbers must you use if you want to compute Keri's standard score?

- A. 30, 70, 100
- B. 30, 70, 200
- C. 60, 70, 100\*
- D. 60, 70, 200

**Analyze: Differentiating**

A rat has been trained to press a bar to receive small pieces of food. The food dispenser has now been disconnected, but the rat is still hungry. Which of the following will have the greatest influence on how long the rat will continue to press the bar?

- A. the reinforcement schedule used during training\*
- B. the total amount of food the rat ate during training
- C. the number of times the rat pressed the bar during training
- D. the amount of energy the rat expended to press the bar during training

**Analyze: Organizing**

You place your finger in the middle of your forehead, and then move it straight back along the middle of your head and down the back of your head. Your finger started out near your frontal lobe and ended up near your cerebellum. Which parts of your brain did it come near along the way?

- A. first the parietal lobes, then the temporal lobes
- B. first the parietal lobes, then the occipital lobes\*
- C. first the temporal lobes, then the parietal lobes
- D. first the temporal lobes, then the occipital lobes

**Analyze: Organizing**

Suppose that you are reviewing the history of the research that has been carried out on a particular topic over a very long period of time. Which of the following patterns would be most likely to characterize the methodological progress of the research?

- A. case studies first, then experimental studies, then correlational studies
- B. case studies first, then correlational studies, then experimental studies\*
- C. experimental studies first, then case studies, then correlational studies
- D. experimental studies first, then correlational studies, then case studies

**Analyze: Attributing**

During a session with a client, a therapist says the following: “The problems that you are having now can be traced back to your relationship with your father when you were a little boy.” Which of these theorists has probably had the greatest influence on this therapist?

- A. Aaron Beck
- B. Carl Rogers
- C. B.F. Skinner
- D. Sigmund Freud\*

**Analyze: Attributing**

Which of the following would a Rogerian therapist be most likely to say during a session with a client?

- A. You seem to be feeling a bit down today.\*
- B. I want you to try to work things out with your sister.
- C. Your dream about going to the zoo—what do you think it might signify?
- D. There are some things I want you to work on before we meet again next week.

**Analyze: Attributing**

Your friend Anne reads that census data indicate that people are having fewer children nowadays than they did 40 years ago. Anne does not believe this because the young couple who live next door to her are both under 30 and already have four children. If Keith Stanovich were told about this, what might you reasonably expect him to say?

- A. The census data must be wrong.
- B. Anne’s comment illustrates valid probabilistic reasoning.
- C. Anne’s comment illustrates the use of person-who statistics.\*
- D. The young couple provide an exception that actually serves to prove the rule.

**Evaluate: Checking**

You have carried out a  $3 \times 2$  ANOVA for independent groups. There were 60 participants, with 10 participants randomly assigned to each cell. You are double-checking your work. Which of the following would immediately let you know that you have made an error?

- A. You found the total degrees of freedom to be 60.\*
- B. You found the mean square for the error term to be 6.25.
- C. You found the  $F$ -statistic for the interaction effect to be 2.34.
- D. You found degrees of freedom for the interaction effect to be 2.

**Evaluate: Checking**

Which of the following research findings would most strongly suggest that differences in Variable X are influenced by genetic factors?

- A. Sisters reared apart have more similar scores on Variable X than sisters reared together.
- B. Sisters reared together have more similar scores on Variable X than sisters reared apart.
- C. Same-sex fraternal twins reared together have more similar scores on Variable X than identical twins reared together.
- D. Identical twins reared together have more similar scores on Variable X than same-sex fraternal twins reared together.\*

**Evaluate: Checking**

Which of the following research findings would most strongly support Schachter's theory of emotion?

- A. People injected with adrenaline find a slapstick movie funnier than those given a placebo.\*
- B. Some people are more expressive than others when they experience strong emotions.
- C. People who suffer damage to the amygdala frequently mistake expressions of anger for smiles.
- D. People who hold a pen in their teeth find cartoons funnier than people who hold a pen in their lips.

**Evaluate: Critiquing**

John is a healthy, well-adjusted young man who happens to become very nervous whenever he has to speak in public. To get over this problem, he starts taking some capsules that his doctor has told him will make him feel calmer when he speaks in public. John does not know it, but the capsules are actually empty. Now suppose that John takes a capsule the next time he speaks in public. How is likely to feel?

- A. more nervous than usual
- B. just about as nervous as usual
- C. less nervous than usual\*
- D. It is not possible to make a reasonable prediction.

**Evaluate: Critiquing**

Bill wants to compare the effectiveness of two training methods for teaching people to juggle. He obtains a group of non-jugglers and randomly assigns each person to one of the two training methods. He sets alpha at 0.05, two-tailed, and he determines that beta is equal to 0.70. Which of the following is a valid criticism of this research study?

- A. Power is too low.\*
- B. The probability of a Type I error is too high.
- C. He should use a one-tailed test.
- D. Participants should select their own training method.

## 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 2004; 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?

## Context-dependent Item Set #1

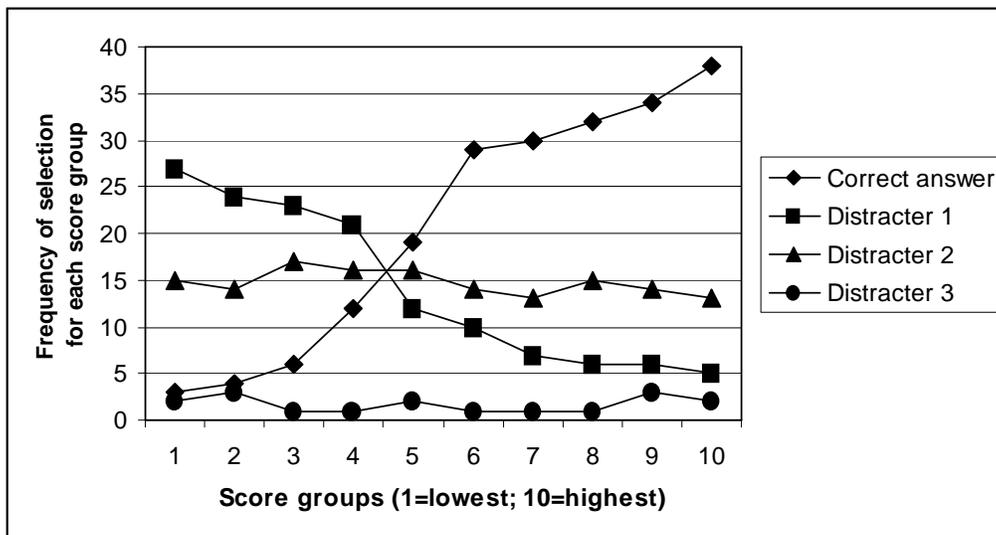
**Instructions:** Read the following research scenario and answer the questions that follow.

*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  $t$ -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 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.

## Context-dependent Item Set #2

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



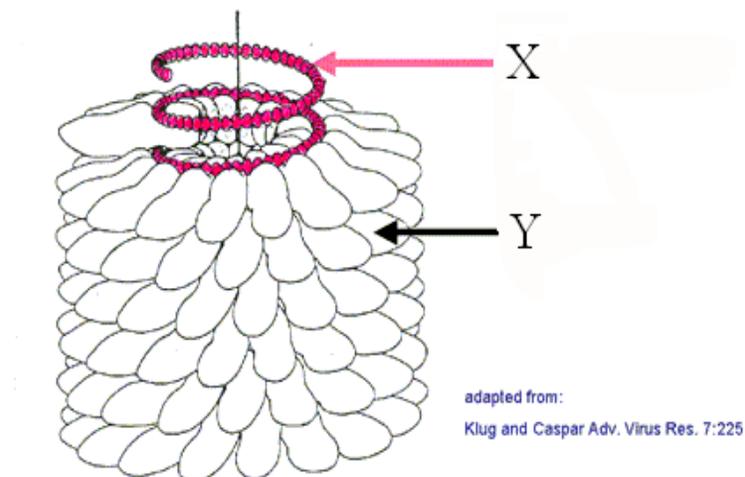
- Which of the following characteristics of the correct answer is most important?
  - It is selected more frequently than any of the distractors.
  - It is selected more frequently than all of the distractors put together.
  - \*High-scoring students are more likely than low-scoring students to choose it.
  - Students with very little knowledge of the topic are likely to answer correctly by guessing.
- Which of the following best characterizes Distracter 1?
  - plausible and a poor discriminator
  - \*plausible and a good discriminator
  - implausible and a poor discriminator
  - implausible and a good discriminator
- Which of the following best characterizes Distracter 2?
  - \*plausible and a poor discriminator
  - plausible and a good discriminator
  - implausible and a poor discriminator
  - implausible and a good discriminator
- Of the following values, which is *closest* to the total number of students that selected Distracter 2?
  - 25
  - 50
  - 100
  - \*150

### Context-dependent Item Set #3: Stimulus only



*Grace before Meat*  
 Jan Steen (c. 1626-1679)  
 (From [www.gardenofpraise.com](http://www.gardenofpraise.com))

### Context-dependent Item Set #4: Stimulus only



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

## Context-dependent Item Set #5

### 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.

*Integer non orci nec nunc tincidunt sodales,  
 Morbi congue nisi a nisl adipiscing tristique.  
 Class aptent taciti sociosqu ad litora torquent,  
 Per conubia nostra, per inceptos himenaeos.*

*Pellentesque et nisl ante donec fringilla  
 Lobortis risus non pharetra risus tempor:  
 At nulla eu risus-nunc id interdum massa  
 Donec non nibh tincidunt nulla.*

*Duis varius placerat felis id laoreet velit,  
 Aliquam nec sed interdum sapien quis ante  
 Facilisis non dignissim, ipsum egetas  
 Suspendisse potenti ut rutrum ullamcorper,  
 Nisi sit amet, vulputate nibh rutrum amet,  
 Cras sollicitudin eleifend ante bibendum.*

Robert Graves (1895-1985)  
 From "Poems and Satires" (Cassell, 1951)

---

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

- The poetic form of "The Cool Web" is best characterized as
  - free verse with a concluding rhymed couplet.
  - ballad stanzas with irregular rhymes.
  - blank verse with unusually irregular rhythm.
  - partially rhymed quatrains with a concluding sestet.
- To emphasize the adult loss of childhood experience, the speaker of the poem
  - relies on frequent breaks in the middle of the line.
  - establishes a tone of caution, nostalgia and forgetfulness.
  - uses imagery of drowning.
  - 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.

and so on...

---

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 Using Context-dependent Item Sets

In a context-dependent item set, a number of multiple-choice items follow the presentation of novel stimulus material, such as a reading, scenario, data set, chart, or map. To answer the multiple-choice items correctly, students must refer to the stimulus material and make use of course-based knowledge and skills. When using context-dependent item sets, keep the following in mind:

- Prepare stimulus material appropriate for the course in which it is to be used.
- Ensure that the stimulus material is **novel**.
- Make stimulus material as brief and as clear as possible.
- When creating multiple-choice items, follow the usual item-writing guidelines.
- Construct multiple-choice items that assess higher-order cognitive processes, not memory or fact-finding skill.
- Let the number of multiple-choice items be proportional to the length of the stimulus material.
- Context-dependent item sets can be used not only with multiple-choice items, but also with constructed-response questions.
- To reuse an item set, retain the form of an earlier item set and insert new stimulus material. Then make whatever minimal changes are necessary to harmonize the multiple-choice items with the new stimulus material.

Adapted from Linn and Gronlund (1995)

## Further Reading on Multiple-Choice Testing

**This short article is a tremendous place to start reading about multiple-choice testing.**

Tarrant, M. & Ware, J. (2012). A framework for improving the quality of multiple-choice assessments. *Nurse Educator*, 37(3), 98-104.

**This journal article examines the guidelines for writing multiple-choice items.**

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.

**This short article is helpful in ensuring that there is good alignment between the course learning objectives and the items on your test.**

Fives, H., and DiDonato-Barnes, N. (2013). Classroom test construction: The power of a table of specifications. *Practical Assessment, Research & Evaluation*, 18(3). Available online at <<http://pareonline.net/getvn.asp?v=18&n=3>>.

**These books focus specifically on multiple-choice testing.**

Haladyna, T.M. (2004). *Developing and Validating Multiple-Choice Test Items*, 3<sup>rd</sup> edition. Mahwah, New Jersey: Lawrence Erlbaum Associates.

McDonald, M.E. (2007). *The Nurse Educator's Guide to Assessing Learning Outcomes*, 2<sup>nd</sup> edition. 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*, 5<sup>th</sup> 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.

**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 are taking a 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(4), 111-131.

IFAT forms can be obtained from Epstein Educational Enterprises. If you wish to purchase forms or would like more information, go to <[www.epsteineducation.com](http://www.epsteineducation.com)>. (Note that I have absolutely no financial interest in the IFAT. I just think that it is a great idea!)

You may reach me at <[david.dibattista@brocku.ca](mailto:david.dibattista@brocku.ca)>.