Writing Effective Multiple-Choice Questions

Bobbie Osborne
Arts and Science Online
What are the benefits and limitations of multiple-choice assessments?
“A well-constructed multiple-choice test can yield test scores at least as reliable as those produced by a constructed-response test, while also allowing for broader coverage of the topics covered in a course.”

(DiBattista, 2011)
A study concluded that perhaps as many as 10 to 15% of medical students failed their exams when they should have passed.

(Downing, 2005)
“While many faculty take seriously their formal lecturing duties and invest considerable thought into the preparation of lectures, ... few commit commensurate effort to the task of preparing examination questions on the material they teach.”

(Jozefowicz et al., 2002)
“All too often (high-stakes exams) are written and assembled at the last minute... The end result is that many excellent teachers assess their students using examinations of questionable quality.”

(Jozefowicz et al., 2002)
Three Challenges of MCQ

Challenge #1
Many multiple-choice questions contain item-writing flaws. (DiBattista, 2011)
Three Challenges of MCQ

Challenge #1
Many multiple-choice questions contain item-writing flaws.

Challenge #2
Most multiple-choice tests will do no more than test factual recall.

(DiBattista, 2011)
Three Challenges of MCQ

Challenge #1
Many multiple-choice questions contain item-writing flaws.

Challenge #2
Most multiple-choice tests will do no more than test factual recall.

Challenge #3
Multiple-choice questions save marking time but require more time to write.

(DiBattista, 2011)
Agenda

• What makes a multiple-choice question a reliable and valid assessment, able to distinguish between high- and low-achieving students
• What are common flaws that we can avoid?
• How can we use multiple-choice assessments to assess higher-order cognitive skills, rather than simply recall?
• What time-saving techniques can I use to write effective questions quickly.
What makes an assessment effective?
What makes an assessment effective?

Good multiple-choice assessments are made up of individual test items that are **valid** and **reliable** in assessing students’ knowledge.
What makes an assessment valid?

Tests with a high level of validity provide an accurate measure of the student’s knowledge and abilities.

- Does the test measure what it’s intended to measure (i.e., whether students have achieved the learning outcomes)?
- Do the test questions accurately reflect the content of the course?
- Does the balance of items reflect the level of emphasis they were given in the course?
- Will a student receive a comparable mark on a different test of the same subject matter?
What makes an assessment reliable?

Tests with a high level of reliability produce **consistent, reproducible measurement**.

- Will students with comparable knowledge and ability receive comparable marks?
- Will the marks given by different TAs be consistent with each other?
- When questions are drawn from a question pool, will each exam produce comparable results?
- If there are multiple versions of an exam, will each version produce comparable results?
Ensure sampling validity.

Exams should be consistent with the activities and assessments used throughout the course.
Ensure sampling validity.

Exams should be consistent with the activities and assessments used throughout the course.

The distribution of questions on an exam should roughly correspond to the level of emphasis each topic received during the course.
Reduce irrelevant difficulties.
Isolate the variable you want to measure.

- Assess knowledge and skills developed in the course, rather than other factors, such as working memory capacity.
- Edit and proof the test items.
- Avoid phrasing question in a way that rewards students for using a process of elimination to arrive at the correct answer.
Guidelines for Writing Effective MCQs
Anatomy of a multiple-choice test item

What is the name of this musical symbol?

a. *fermata  
b. caesura  
c. breve  
d. tenuto

Stem

Keyed answer

Distractors

Answer options
Writing the stem

1. The stem should present a single, clearly-defined problem.
2. The stem should be able to stand on its own as a short answer question.
3. Minimize the amount of reading required.
4. Use simple, clear language to avoid misunderstandings.
Writing Distractors

All answer options should be plausible.
To generate plausible distractors,
• Use statements that are true, but do not answer the question posed.
• Use the common errors that students make.
• Use words that sound important or have associations with the stem.
Distractors should be plausible.

Which of the following is an example of a summative assessment?

*a. Midterm exam
b. Golf

Each answer option has a 1 in 4 chance of being correct.
c. Tetris
d. Peer review
Distractors should be plausible.

Which of the following is an example of a summative assessment?

*a. Midterm exam*

*b. Golf*  

c. Tetris  

d. Peer review

Each answer option has a 50:50 chance of being correct.
Writing Distractors

All answer options should have the same domain.

In the final scene of *Hamlet*, which character fights Hamlet in a duel?

a. Petrucchio  
b. Tybalt  
c. Banquo  
d. *Laertes
Writing Distractors

All answer options should have the same domain.

In the final scene of *Hamlet*, which character fights Hamlet in a duel?

a. Petrucchio (*The Taming of the Shrew*)
b. Tybalt (*Romeo and Juliet*)
c. Banquo (*MacBeth*)
d. *Laertes (*Hamlet*)*
In the final scene of *Hamlet*, which character fights Hamlet in a duel?

a. Horatio (*Hamlet*)
b. Fortinbras (*Hamlet*)
c. Guildenstern (*Hamlet*)
d. Laertes (*Hamlet*)
Dizygotic twins ...

a. have identical DNA.
b. appear to be identical.
c. rarely resemble each other.
*d. grow from separate zygotes.
The stem should present a clearly-defined problem.

Avoid grouping together multiple true-or-false statements.

Which of the following statements is TRUE?

a. Ireland was colonized by the ancient Romans.
b. Columbus was the first European to sail to the Americas.
c. The Incas had written language prior to European contact.
*d. Documents written in Classical Greek don’t use spaces between words.
The stem should present a clearly-defined problem.

Avoid grouping together multiple true-or-false statements.

Which of the following is a myth about psychology?

a. Experience physically alters the structure of the brain.
b. We are consciously aware of only a small fraction of our mental activity.
*c. Heritable traits, such as IQ, cannot be altered by experience.
d. The mind is a product of a physical machine, the brain.
Ask a question.

Students _______ would be the most representative sample for a study of competitiveness in high school.

a. competing in sports
b. *in a required health class
c. in advanced math classes
d. who agree to do the study during lunch
Slightly Revised
The most representative sample for a study of competitiveness in high school would be students _____

a. competing in sports.

b. *in a required health class.

c. in advanced math classes.

d. who agree to do the study during lunch.
Revised

Which of these groups would be the most representative sample for a study of competitiveness in high school?

a. Students on a sports team
*b. Students in a required health class

c. Students in advanced math classes
d. Students who agree to do the study during lunch
The photoreceptors we need to appreciate a blue sky on a sunny day are the _______, whereas the photoreceptors we use to detect the stars outside at night are the _______.

a. cones; hair cells
b. Meissner’s corpuscles; rods
*c. cones; rods
d. rods; cones
In cases where only one of two identical twins develops schizophrenia, all of the following factors may be responsible except

a. birth weight
* b. birth order
  c. nutrition
  d. nurturing
Which of the following is a term used to describe twins who do not share 100 percent of their genes?

*a. dizygotic twins
b. monozygotic twins
c. identical twins
d. discordant twins
Revised

Which of the following is a term used to describe twins who do NOT share 100 percent of their genes?

*a. dizygotic twins
b. monozygotic twins
c. identical twins
d. discordant twins
In what year were Canadian women (except for Indigenous women) first eligible to vote in federal elections?

a. 1916
b. 1921
* c. 1918
b. 1917
Revised

In what year were Canadian women (except for Indigenous women) first eligible to vote in federal elections?

a. 1916
b. 1917
* c. 1918
b. 1921
Which of the following SSRIs has an active metabolite AND the longest half-life?

a. Citalopram  
b. Sertraline  
c. Paroxetine  
d. Fluoxetine
Who received a Nobel prize for the discovery of the structure of DNA?

1. Francis Crick
2. James Watson
3. Rosalind Franklin

*a. 1 and 2
b. 2 and 3
c. 1 and 3
Who received a Nobel prize for the discovery of the structure of DNA?

1. Francis Crick
2. James Watson
3. Rosalind Franklin

*a. 1 and 2
b. 2 and 3
c. 1 and 3*
Reward knowledge and skill, not pattern recognition.
Which of the following medicines is used to treat viral infections?

a. acetaminophen  
b. phenylephrine  
c. ibuprofin  
*d. penicillin EXCEPT if the patient is known to have an allergy to penicillin
The longest or most complex option is usually correct.

What is the best reason for listing information sources in your research assignment?

a. It is required
b. To get a better grade
c. To make it longer
d. It is unfair and illegal to use someone's ideas without giving proper credit*
The option using absolute terms is usually incorrect.

Which of the following medicines is always used to treat viral infections?

a. acetaminophen
b. phenylephrine

c. ibuprofen
*d. penicillin
The option that repeats words in the stem is usually correct.

According to the Big Bang theory, what event occurred to produce the universe?

a. a meteor shower
b. a contraction and expansion of matter
c. current flowing along plasma filaments
d. current flowing in a galactic magnetic field
e. an explosion
Which of the following groups has the highest correlation of IQ scores between family members?

A. identical twins reared apart
B. identical twins reared together
C. fraternal twins
D. siblings who aren’t twins
One of the paired options is usually correct.

Revised
Which of the following groups has the highest correlation of IQ scores between family members?

A. identical twins reared apart
B. identical twins reared together
C. fraternal twins reared apart
D. fraternal twins reared together
According to the Big Bang theory, how did the universe begin?

a. with a meteor shower
b. the universe is in a steady state
c. the universe expands and contracts
d. with electric currents flowing along plasma filaments
e. with current flowing in a galactic magnetic field
d. *with an explosion

Options with grammatical errors are incorrect.
All of the above is usually correct.

Which of the following are measures of central tendency?

a. Mean
b. Median
c. Mode
*d. All of the above
None of the above is usually incorrect.

Who is the author of *Middlemarch: A Study of Provincial Life*?

a. Charlotte Brontë

b. Thomas Hardy

c. Charles Dickens

*e. None of the above*
None of the above is usually incorrect.

Who is the author of *Middlemarch: A Study of Provincial Life*?

a. Charlotte Brontë  
b. Thomas Hardy  
c. Charles Dickens  
*e. None of the above  George Eliot*
When in doubt, guess “C” (if the quiz is analog).
Phrase questions to limit the possible answers.

Which of the following...
Close loopholes.

How would you describe ...
Why do you think...
Three Challenges of MCQ

Challenge #1
Many multiple-choice questions contain item-writing flaws.

Challenge #2
Most multiple-choice tests will do no more than test factual recall.

Challenge #3
Multiple-choice questions save marking time but require more time to write.

(DiBattista, 2011)
Three Challenges of MCQ

Challenge #2
Most multiple-choice tests will do no more than test factual recall.

- Use novel material, rather than examples presented in the course.
- Ask students to apply what they’ve learned in a novel context.

(DiBattista, 2011)
Assess terminology through application.

In classical conditioning, what term is given to a stimulus that elicits a particular response even in the absence of any prior training?

a. conditioned stimulus
b. unconditioned stimulus*
c. activational stimulus
d. discriminative stimulus
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. unconditioned stimulus* 
C. activational stimulus  
D. discriminative stimulus
Getting beyond remembering.
Use a test blueprint

<table>
<thead>
<tr>
<th>Bloom's Taxonomy</th>
<th>Topic or Learning Outcome 1</th>
<th>Topic or Learning Outcome 2</th>
<th>Topic or Learning Outcome 3</th>
<th>Topic or Learning Outcome 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remember</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>5 (14%)</td>
</tr>
<tr>
<td>Understand</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>7 (17.5%)</td>
</tr>
<tr>
<td>Apply</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>15 (37.5%)</td>
</tr>
<tr>
<td>Analyze</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>10 (25%)</td>
</tr>
<tr>
<td>Evaluate</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>2 (5%)</td>
</tr>
<tr>
<td>Total</td>
<td>10 (25%)</td>
<td>10 (25%)</td>
<td>10 (25%)</td>
<td>10 (25%)</td>
<td>40</td>
</tr>
</tbody>
</table>
Ask students to apply knowledge and skills in novel contexts.

- Which of the following best defines $x$?
- What is the meaning of $x$?
- What is a defining characteristic of $x$?
- Which of the following is an example of $x$ ... ?
- What approach would you use to...?
- Which statement best exemplifies the principle of $x$?
- What would result if...?
Ask students to identify motives or causes, make inferences, find evidence.

- What is the cause (or reason for) x?
- How is x related to y?
- What is the relationship between x and y?
- What distinguishes x from y?
- Which of the following is the most (or least) important (or significant, effective, etc.)?
- How could you determine ....?
- How could you measure (test)....?
Question Shells: Evaluate

Ask students to make predictions or propose alternatives.
• What would happen if ...?
• What changes would you make to ...?
• What could be done to minimize/maximize, increase/decrease, etc....?
• Can you propose an alternative to ...?
• Can you predict the outcome if...?
• How would you prove (or disprove) ...?
• If x occurs, what will most likely be the result?
Create a context-dependent item set.

• Select an appropriate image, scenario, chart, graph, data set, reading, etc.
• Write a series of questions that ask students to use what they have learned to analyze the stimulus material.
Thank you!