Writing Effective Multiple-Choice Questions

What are the benefits of multiple-choice testing?

- Efficient – Tests a broader range of material in a shorter amount of time.
- Reliable – Scores are highly consistent between markers.
- Scalable – Can be used with any class size.

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).

What are the challenges of using multiple-choice questions to assess students?

Challenge 1: Many multiple-choice test items are contain item-writing flaws.
Challenge 2: Most multiple-choice test items are used to test factual recall.
Challenge 3: Multiple-choice test items save marking time but require more time to write (DiBattista, 2011).

A trade-off: Multiple-choice exams are quick to mark, and the results of the assessment are more reliable, BUT more time needs to be invested up-front to write questions that provide effective assessments of students’ knowledge and skill.

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.

To ensure sampling validity,

- Align questions with the learning outcomes of the course.
- 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.

Tips

1. Write the exam questions as you prepare the content for the course.
   - “We often caution students to avoid cramming for an exam. Likewise, we’d like to encourage faculty to spend some time each week writing exam or quiz items to build a pool of questions that they can use on the exam” (Towns, 2014).
2. List the learning outcomes for the course and use this as an outline for your exam.
   - Each learning outcome should be assessed.
   - Each test item should assess a single learning outcome.
3. Decide on an appropriate weighting for each learning outcome.
4. Write questions.

Effective assessments isolate the dependent variable: students’ knowledge and skill at achieving the intended learning outcomes.

Item-writing flaws tend to reward students who
   - Have higher capacity working memories,
   - Have above-average reading proficiency,
   - Are native English speakers, and/or
   - Use test-taking strategies to receive full marks for partial knowledge.

**Don’t reward English proficiency if it is not related to the learning outcomes or subject matter.**
   - Any confusion over grammar or question structure invalidates the test.
   - In most disciplines, knowledge of grammar is not related to the learning outcomes or subject matter.
   - Use correct grammar, punctuation, capitalization, and spelling.
   - Use vocabulary that is appropriate for the course level.
   - Avoid the use of idioms.
   - Edit and proof items.

**Reward knowledge and skill; rather than test-taking strategies.**
   - Many multiple-choice questions are worded in such a way that they give full marks for partial knowledge.

**The Anatomy of a Multiple-Choice Question**
A multiple-choice question has three components:
   - the question stem,
   - the keyed response (correct answer option), and
   - the distractors (incorrect answer options)

When did Canadian women (except Indigenous women) become eligible to vote in federal elections?

a. 1915  
   b. 1916  
   c. 1917  
   d. *1918

Stem  
Distractors  
Keyed answer  
Answer options
Guidelines for writing effective Multiple-Choice Questions
There are well-established, evidence-informed guidelines for writing effective multiple-choice questions.

Writing a better stem.
The stem should present a single, clearly-defined problem.

The stem should be able to stand on its own as a short answer question.

- All the relevant information required to answer the question should be placed in the stem, rather than in the answer options.
- If you covered up the answer options, would students have all of the information that they would need to provide a written answer?

Minimize the amount of reading required.

- The stem should only include the information needed to make the problem clear and specific. Extraneous information that is not required for the assessment increases reading time and decreases the number of questions that can reasonably be answered within the timeframe and hence, the reliability of the test.
- Keep the answer options as short as possible.

Use simple, clear language to avoid misunderstandings.

Writing better answer options.

All answer options should be plausible.

- Three distractors are adequate. Writing more than three distractors has little impact because the additional distractors are likely to be less plausible.
- To generate plausible distractors,
  - Use statements that are true, but do not answer the question.
  - Use common errors.
  - Use words that sound important or have associations to the stem.

Distractors should share the same domain.

All answer options should be fairly equal in length.

Item-Writing flaws that increase irrelevant difficulty

The stem should pose a clearly-defined problem; avoid unfocused stems.

- Students should know what is expected of them after reading the stem; test items shouldn’t begin with a sentence stub.
- These questions likely combine more than one concept or learning objective and test recall.
Items that ask “Which of the following is true (or false)…” do not pose a clearly-defined problem; they simply group a number of true or false statements together. There is no advantage to this.

**Example of sentence stub**
Dizygotic twins …
a. have identical DNA.
b. appear to be identical.
c. rarely resemble each other.
*d. grow from separate zygotes.*

**Ask a question. Use fill-in-the-blanks only when necessary.**
- The exception: A short, straightforward command is fine, e.g. Calculate… Derive the formula… Solve for …

If you do use fill-in-the blank format (a.k.a. sentence completion format),
- The answer blank should appear at the end of the test item, rather than in the middle.
- Don’t use double-barrelled fill-in-the-blanks.

**Avoid negative wording, whenever possible.**
- Negative wording makes the question unnecessarily difficult as students have to mentally switch from finding the correct answer to finding incorrect ones (Boland, Lester, & Williams, 2010).
- Students may be nervous when they write an exam and they can easily skip over words such as not or except and misconstrue the question.
- If you use negative wording, emphasize negative words by using ALL CAPS or boldface.

**Use formatting to enhance clarity.**
- Format test items vertically instead of horizontally.
- Arrange answer options in a logical order (chronological, numerical, etc.). (In onQ you can turn off scrambling answer options for specific questions.)

**Ask only one question at a time; avoid asking double-barrelled questions.**
Which of the following SSRIs has an active metabolite AND the longest half-life?

**Keep answer options independent of each other; avoid grouping answer options together to create a Type K question.**
**Example of Type K question**
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
• This type of question involves irrelevant difficulty as students must carefully read through the options to distinguish between the choices and makes large demands on working memory (cognitive overload).
• The argument is sometimes made that Type K items test higher-order thinking skills; however, research data suggests that this is not the case.
• They reward full marks for partial knowledge.
• Students use a process of elimination to guess the correct answer.
  o Students are three times more likely to guess the right answer (Albanese, 1993)
  o Are 40% less reliable (Albanese, Kent, and Whitney, 1979)

**Item-Writing flaws that reward pattern recognition**

**The answer that is the longest or most complex is usually the correct answer option.**
- Why? To ensure that an answer option is unequivocally correct, test writers add additional details and caveats which leads to the response being overly long.

**An answer option that includes absolute terms is usually not the correct answer.**
- Be cautious about using terms such as *always/never, all/none, completely*, etc. “always” or “never” in answer options because they are overly specific and usually indicate that the option is wrong.

**The answer option that repeats words or phrases (or includes closely related words) in the stem is usually the correct answer.**

**If specific terminology (jargon) is used in one of the distractors, it is usually the correct answer.**

**When the options include a single pair of opposites, the keyed option is almost always a member of the pair.**

**If an answer option contains spelling or grammatical errors or doesn’t follow grammatically from the stem, it is not the correct option.**
- Item writers tend to pay more attention to the correct answer than to the distractors. So, when a test item contains a grammatical error, the answer option that contains the error is usually wrong.
- Use parallel sentence construction.
- Ensure answer options are in the same category (all theories, all terms, all procedures, all names, etc.)
- Pay particular attention to ensure verb tenses agree with nouns (singular or plural) and that the appropriate indefinite article is used before nouns that begin with a vowel.
If “all of the above/all of the options” is an answer option, it’s usually the correct answer.
- All of the above/all of the options converts a multiple-choice question into a series of true/false questions.
- If a test taker recognizes that more than two answer options are correct, he or she will conclude that all must be correct.
- you will never know if students know the correct answer.

If “none of the above/none of the options” is an answer option, it’s usually NOT the correct answer.
- None of the above/none of the options converts a multiple-choice question into a series of true/false questions.
- Students will receive full marks, even if they may not know the correct answer.

When in doubt, guess option C (if the quiz is analog).
- Test writers seem reluctant to have the right answer appear first or last.
  Solution: If you’re using onQ, scramble the order of answer options.

Also, randomizing the question order will make the test more reliable.
- Students can be primed to perform better on a test when questions are massed by topic or learning outcome. Weaker students may do better on the test than they would otherwise.

Student-proof your tests.
Make sure that only one of the answer options is the right answer.

Phrase questions to limit the answer to the answer options listed to preclude arguments.
- Which of the following …
- Which of the following is the best answer?
- Which of the following is most correct?

Close loopholes.
- How would you describe (explain) …?
  Why do you think …?
  Students can argue that their answer does reflect how they would describe something or what they think about something.
Getting beyond remembering.
Using Bloom’s taxonomy one can classify items as requiring lower-level cognitive skills (remember, understand) or higher-order cognitive skills (apply, analyze, evaluate).

A test blueprint can be used to identify the categories to be tested and establish targets for the relative weight given to each. A reasonable target would be that 30% to 50% of items on a test or exam assess higher-order cognitive skills.

<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 1</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</td>
</tr>
<tr>
<td>Understand</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Apply</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>Analyze</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Evaluate</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>2</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>

Use novel material to test higher-order thinking.
- If you use an example exactly as it was presented in the textbook, you’re testing whether students can recall what they read, rather than whether they understand or can apply concepts in novel contexts.
Ask students to apply a definition in a new situation.

- To create application questions that test definitions, place the concept in a life situation or context that requires the student to first recall the facts and then apply or transfer the application of those facts into a situation.

Tips for writing questions

Use Item Shells

Haladyna and Shindoll (1989) recommend using item shells to construct multiple-choice questions. Item shells can be used like algorithms to vary the specifics.

For instance,

[[Name]], a [[#]-year-old [[gender]] patient is brought to the [[hospital, clinic, etc.]], complaining of [[insert symptoms]]. She has a history of [[medical condition]] that has previously been treated with [[treatment]]. Which of the following [[treatment, intervention, laboratory studies]] are indicated?

The following item shells are adapted from Haladyna (1997 and 2004).

Understand (Explain ideas or concepts)
Which of the following best defines x?
What is the meaning of x?
What is a defining characteristic of x?

Apply (Use knowledge and skills in new situations)
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…?

Analyze (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 measure (test)…?
How could you determine ….?

Evaluate (Make predictions; 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 context-dependent item sets
- Present novel stimulus material, such as a picture, reading, scenario, chart, graph, data set, map, etc.
- Create a series of multiple-choice questions that ask students to use the skills that they have developed in the course to analyze the stimulus material.

References
Boland, R.J., Lester, N.A. & Williams, E. Writing Multiple-Choice Questions. Acad Psychiatry 34, 310–316 (2010). https://doi-org.proxy.queensu.ca/10.1176/appi.ap.34.4.310


