

Designing Quality Multiple Choice Questions

Yale Center for Teaching and Learning

<https://ctl.yale.edu> / <https://ctl.yale.edu/MultipleChoiceQuestions>

Multiple choice questions effectively assess student learning because they are flexible, relatively easy to implement and grade, and allow instructors to sample a range of course materials. Additionally, instructors often have access to test bank questions through their departments, previous courses, or online resources. Research suggests that practicing multiple choice questions outside of class - testing recall, conceptual awareness, and even problem-solving - can be as effective as active learning for deeper conceptual gain (Karpicke and Blunt, 2011). In designing or choosing multiple choice questions for assessments, instructors can take steps to ensure that questions are accurate and reliable measures of student achievement.

Reliability depends largely on grading procedures and question format. Multiple choice questions are typically composed of (1) a question *stem* and (2) several choices, including *distractors* and *one correct option*. Research literature suggests including a minimum of 3 answer choices (Haladyna et al. 2002).

- *Good stems* introduce the central idea clearly and briefly, and are constructed in positive ways that avoid words such as “not” or “except” (so, “In the Dred Scott decision, the Supreme Court enabled slaveowners to ...”, instead of “In the Dred Scott decision, the Supreme Court’s decision did not ...”).
- *Good distractors* target major student misconceptions, remain clear by avoiding overlapping content and awkward grammar, and avoid catch-all answers like “all of the above” or “none of the above.” Additionally, effective distractors should not provide clues to rule them out, and should shuffle order with the right answer from question to question.

Instructors can consider a variety of examples and recommendations for constructing effective multiple choice (MC) questions.

Examples

In the following examples of effective and ineffective MC questions, students explore potential energy, or the energy that is stored by an object.

#1. Good Stem, Poor Distractors

Potential energy is:

- a) the energy of motion of an object.
- b) not the energy stored by an object.
- c) **the energy stored by an object.**
- d) not the energy of motion of an object.

In this question the good stem is clear, brief, and presents the central idea of the question through positive construction. However, the distractors are confusing: b) and d) are written in negative constructions that force students to reinterpret the stem, while c) and d) have overlapping, inconsistent content that confuses and tests reading comprehension over content recall. Finally,

choices do not move logically by grouping content, failing to visualize and test larger concepts for students.

#2. Poor Stem, Good Distractors

Potential energy is not the energy:

- a) **of motion of a particular object.**
- b) stored by a particular object.
- c) relative to the position of another object.
- d) capable of being converted to kinetic energy.

In this question the poor stem contains the word “not,” which fails to identify what potential energy *is*, and tests grammar over student understanding. However, the good distractors are written clearly, cover unique content, and follow a logical and consistent grammatical pattern.

#3. Good Stem, Good Distractors

Potential energy is:

- a) the energy of motion of an object.
- b) **the energy stored by an object.**
- c) the energy emitted by an object.

In this example both the stem and the distractors are written well, remain consistent, and test a clear idea.

Recommendations

- **Align MC questions with course learning objectives and class activities** - Effective MC questions tie back to class activities and learning goals, so that students can apply their knowledge and see connections throughout their learning. Instructors can use backward design to first write course learning outcomes, and then activities and assessments that work towards those outcomes. Among the assessments, MC questions can test recall, conceptual knowledge, and problem-solving.
- **Use Bloom’s Taxonomy to develop MC questions that assess different levels of cognition** - Bloom’s Taxonomy, revised in 2001, categorizes six levels of learning from lower- to higher-level thinking skills: Remember, Understand, Apply, Analyze, Evaluate, and Create. Ideally, learning outcomes, class activities, and assessments map to these ascending levels throughout a class or course. Instructors can develop or select questions that assess different levels of cognition, keeping in mind that MC questions at the Create level may prove more challenging to develop.
- **Provide opportunities for students to practice taking MC questions before exams** - It remains a common misconception that students learn optimally just by reviewing their course notes and/or reading from textbooks. Research shows that more active cognitive involvement, paired with retrieval practices, leads to better student learning. Instructors can

use MC questions to support this kind of learning. Low-stakes MC questions in class allow students to practice material as they prepare for a test, and these questions can be approached individually, through group work, class discussion, case study, or a variety of other means; moreover, instructors and students can use these practice outcomes to modify instruction and study habits. Additionally, instructors can encourage students to take practice multiple choice questions aligned with course material outside of class while studying for tests.

- **Analyze tests to determine which questions do not perform well** - Many institutions have statistical software programs that provide reports on the quality of multiple choice questions. Two common reliability tests are item difficulty and item discrimination. Item difficulty is typically reported as the percentage of those taking the test who chose the correct answer for an item. In classical testing theory (McCowan and McCowan, 1999), the optimal difficulty guideline for a 4-option multiple choice test is 63%. Item discrimination is the measure of how an item detects differences between higher and lower scores on a test. Typically, this is measured by the point-biserial, which is a correlation coefficient. Items that discriminate well are answered correctly more often by the higher scoring students and have a higher positive correlation. Good discrimination indexes are 0.4 and above, poor are 0.2 and below. Instructors can use this data from the test to remove or revise lower quality MC questions.
- **Utilize Canvas** - Yale's learning management system is Canvas, which features a variety of tools for constructing MC questions, grading anonymously, and reviewing statistics for outcomes.
- **Accessibility Awareness** - Instructors utilizing MC questions should be aware of student accessibility concerns, and provide dynamic policies to support students who are print-disabled or may need more time during testing.

References

Haladyna TM, Downing SM, Rodriguez MC. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education* 15(3): 309-334.

Karpicke JD and Blunt J. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science* 331: 772-775.

McCowan RJ and McCowan SC. (1999). Item Analysis for Criterion-Referenced Tests. Retrieved from: <https://files.eric.ed.gov/fulltext/ED501716.pdf>