

## Guidelines for writing multiple-choice questions

Multiple-choice questions are much more flexible than True/False questions, even if in some respects, they simply offer more options to choose from. Carefully written, they can be used to measure learning at all levels of Bloom's taxonomy of learning outcomes, even if they are not usually the best choice of assessment for the highest levels of learning.

All objective questions, those that ask for short answers that can be easily judged, are to some degree limited in measuring skills because they do not provide a realistic context. But they can be very useful to measure the learning that forms the basis of proper execution of tasks.

Following are examples of multiple-choice questions that measure various learning levels. But first, let's examine some characteristics of well written multiple-choice questions. Good questions are always difficult to write, and we can underestimate the time it takes.

### Effective multiple-choice questions

The following are characteristics of effective multiple-choice questions

- The question or statement at the top of a multiple choice question (the "stem") should be as clear and unambiguous as possible. For example, if you ask, "Which of these things is the best ... ?" You must specify the context and criteria of what determines the concept "best." There should be no doubt about what you are asking.
- A correct choice should be unambiguously correct in all situations. Especially if you write a question with a single correct answer, make sure it is clear and defensible. Experts should be able to agree. Reread your question in comparison to each choice to be sure they are unambiguously correct or incorrect.
- Unless it is important for the content, the question should not include negative words like "Which of these things is NOT..." or "What is the LEAST..." These forms of questions increase the risk of someone misreading the question, or they could make the correct answer more obvious.
- Even incorrect choices must be plausible and not too easily ruled out, if you want to measure learning. Make each choice attractive.
- All choices should be about the same length. Avoid adding too many details to the right choice, which will make it stand out. It is tempting to do this because it avoids ambiguity about its correctness, but it is better to instead reduce ambiguity in the question stem.
- The correct answer should change positions in the list of choices (a, b, c, etc.) from one question to the next or students will begin to guess this position. It is common to choose B or C as the correct choice because we think of the choices in this order. Writing plausible incorrect answers is difficult, so while we might think of one good one for A, we wait until C, D or E to create the rest.
- Images can be included in multiple-choice questions to enhance their complexity and call for authentic application of skills.

### Examples of multiple-choice questions

1. Examine this question and decide which level of learning outcomes from Bloom's Taxonomy it is attempting to measure:

**TAF DRRN 111600Z 1118/1224 22008KT 4000 DU SCT030 FEW040CB  
TEMPO 1118/1124 06025G35KT 0800 BLDU TSRA  
BECMG 1200/1202 6000 NSC  
PROB30 TEMPO 1214/1220 12025G35KT 2500 TSRA=**

©The COMET Program

Based on the TAF provided by the N'Djamena aviation forecaster, what conditions can the pilot expect to encounter when landing in Niamey at 23:30 UTC? Select all that apply. (Select all that apply.)

- a) Unrestricted conditions with a few scattered clouds
- b) Dust, which will reduce visibility
- c) Temporary conditions of significantly reduced visibility due to high winds, blowing dust, and thunderstorms
- d) Heavy rain and gusty winds, which will reduce visibility
- e) Fog and calm winds, which will reduce visibility

Which level of learning is the highest level being assessed?

- a) Recall or Knowledge
- b) Comprehension
- c) Application
- d) Analysis
- e) Evaluation
- f) Synthesis

-----  
If you decided Application, you are probably correct. The question is trying to determine if the learners can interpret a number of codes within the TAF. It is not asking for an Analysis or Evaluation of the TAF.

2. Now consider the question below. Which is the highest Bloom's Taxonomy level of learning outcome being measured?

Based on the satellite imagery and NWP products you previously viewed (*assume these were already provided*), which phenomena will impact the flight from Chad to Niamey that's expected to land at 23:30 UTC? (Select all that apply, then click Done.)

- a) Dense fog
- b) Severe icing
- c) Frequent thunderstorms
- d) Isolated thunderstorms

Select the *highest* level:

- a) Recall or Knowledge
- b) Comprehension
- c) Application
- d) Analysis
- e) Evaluation
- f) Synthesis

-----

If you had a difficult time with this question, that is understandable. The best answer is probably Synthesis, but it is not clear. This question is attempting to measure whether someone can integrate knowledge to make a complex, critical decision, which can be seen as an example of Synthesis. It includes Analysis of forecasting products and Evaluation of their validity. Of course, it also requires a lot of background knowledge and understanding, and is certainly an example of Application.

This question is a good example of how Bloom's Taxonomy begins to fail in distinguishing learning outcome levels. Synthesis is usually described as creating a new application or product, and a single forecast does not fit this description. However, the question is asking for much more than simple Application. Our conclusion might be that either there is a missing level or that there are many levels of Application, and that the additional Bloom levels are problematic in the taxonomy.

The take-away point, however, is that multiple-choice questions such as this one are measuring high levels of learning outcomes.