Guidelines for Developing Quality Multiple-Choice Questions

Table of Contents

Item Writing Primer

Item Topic Selection

The first step in developing an item is to select a topic that fits the module content outline.

Items should reflect *important* concerns in practice. Item topics should be neither trivial nor esoteric, and they should not require the examinees to make value judgments.
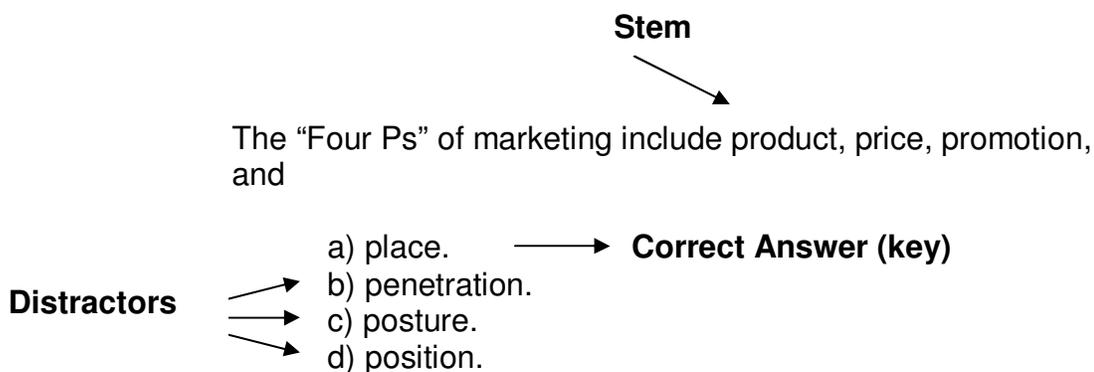
When possible, writers should be able to cite a reference (book, industry, magazine, etc.) that supports the items' correct response. Note that the *actual* answer does not have to be in the reference; rather, the concept should be explained in it.

Item Construction

A multiple-choice question consists of a stem in the form of a question or incomplete statement and alternatives (possible answers), including

- one correct answer (that content experts consider most appropriate)
- distractors – other (typically three) incorrect, yet plausible responses

The example below shows the construction elements.

Stem

The "Four Ps" of marketing include product, price, promotion, and

a) place. **Correct Answer (key)**

Distractors
b) penetration.
c) posture.
d) position.

The above item is in the incomplete statement format. Below is an example of an item in the question format:

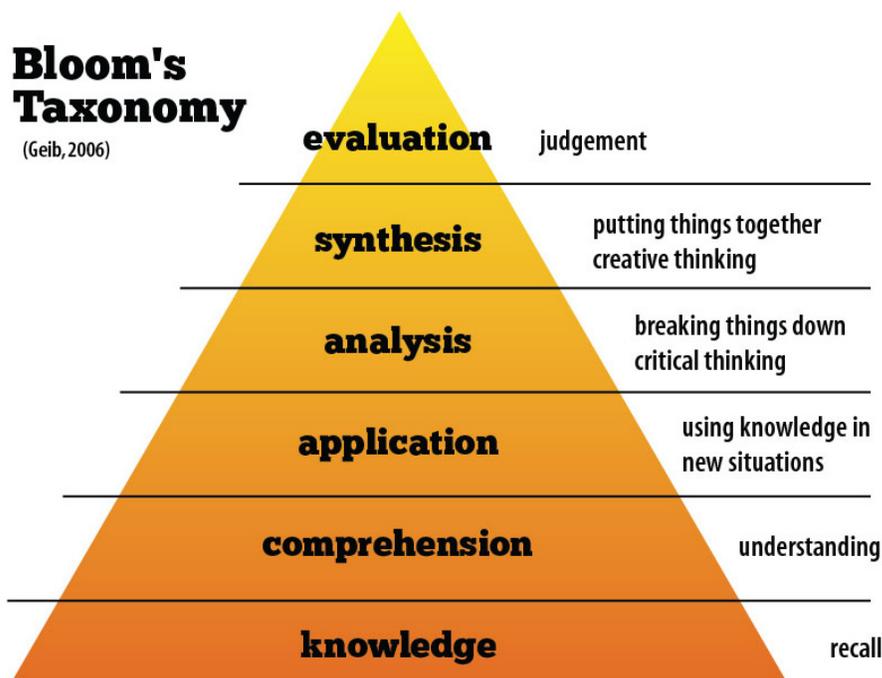Which of the following listings include the "Four Ps" of marketing?
a) place, price, promotion, product
b) place, position, promotion, product
c) penetration, price, promotion, product
d) position, posture, price, product

Both the question and incomplete statement formats are equally acceptable to use.

Item Level

According to Bloom's Taxonomy there are five possible cognitive levels of items: knowledge, comprehension, application, analysis and synthesis. The higher in the taxonomy, the more complex the cognitive level.

Bloom's Taxonomy
(Geib, 2006)

Level	Description
evaluation	judgement
synthesis	putting things together creative thinking
analysis	breaking things down critical thinking
application	using knowledge in new situations
comprehension	understanding
knowledge	recall

While it is easiest to write items at the lower levels that require test-takers to simply recall definitions and basic concepts, in reality, professionals are required to apply their knowledge.

Following are examples of questions at the varying levels.

Knowledge
What is TPN?

Comprehension
Which of the following is an example of TPN?

Application or Beyond
Given a situation, what would be the most important step in initiating TPN?
Given a situation, which macronutrient composition would be most appropriate?

The following table provides definitions of each level and then the key verbs and examples of tasks expected to be demonstrated at this level. This table should serve as a critical resource as you write items. You are being asked to try to limit questions at the knowledge level and begin to focus more on questions at the comprehension and application (or beyond) levels.

Competency Levels	Key Verbs (what competency level is expected at this competency level) and Examples
Knowledge: Recall data or information without modification.	**Key Verbs**: defines, describes, identifies, knows, labels, lists, matches, names, outlines, recalls, recognizes, reproduces, selects, states. **Examples**: Recite a policy. Quote an author from memory. Match a term to its definition.

Comprehension: Understand the meaning, translation and interpretation of material, instructions and problems.	**Key Verbs**: comprehends, converts, defends, distinguishes, estimates, explains, extends, generalizes, gives an example, infers, interprets, paraphrases, predicts, rewrites, summarizes, translates. **Examples**: Summarize a principle. Translate an equation into a computer spreadsheet. State a problem or steps in a process in your own words. Identifies an example of a concept, principle or method. Interprets the meaning of a statement.
Application: Use a concept, method, principle in a new or changing situation.	**Key Verbs**: applies, changes, classifies, computes, constructs, demonstrates, discovers, manipulates, modifies, operates, predicts, prepares, produces, relates, shows, solves, uses. **Examples**: Use a formula to calculate something. Apply a policy in a given situation. Apply a concept, method or principle to identify the most appropriate action in a given situation. Construct a graph or chart of given data.
Analysis: break down information into its parts and understand its organization structure, the relationship among the parts and the principles involved.	**Key Verbs**: analyzes, breaks down, compares, contrasts, correlates, diagrams, deconstructs, differentiates, discriminates, distinguishes, identifies, illustrates, infers, prioritizes, relates, selects, separates. **Examples**: Troubleshoot a problem using logical deduction. Recognize flaws in reasoning. Distinguish between facts and inferences. Compares two methods and selects the appropriate one. Compares options and selects the best one for a given situation. Identifies the cause of a problem and predicts what would have happened if a different action had been taken. Sees correlations. Sees patterns in behavior or data. Draws conclusions from data. Recognizes unstated assumptions. Evaluates the relevancy of data to a particular situation. Recognizes underlying themes.

Item Writing Guidelines

While understanding the cognitive levels provides you guidance on the level of complexity of the questions you write, there are also guidelines you should follow to create quality items regardless of their competency level.

Following are those item writing guidelines that should be followed in writing multiple-choice test questions. They are separated into three sections: overall item guidelines, guidelines associated with item stems, and guidelines associated with item alternatives.

Overall Item Writing Guidelines

1. **Item is linked to the module's content outline.**

2. **Issue addressed by the item is linked to industry reference material, when possible.**

3. **Item does not use language taken verbatim from reference materials.**

4. **Content covered by the item reflects current practice and information, but material is not so new that it is controversial or untested in practice.**

5. **Item presents a realistic situation, not the rare occurrence that could happen, but would be unlikely.**

6. **Item presents universal practices.**
   Practices covered within a certification exam question should be only those that are uniformly practiced across the country – not just a local or regional practice.

7. **Item does not appear to be biased.**
   Items should not discriminate by age, socioeconomic status, gender, ethnicity, religion, culture, geography, or institutional differences OR negatively stereotype population groups.

8. **Item does not include indefinite qualifiers (e.g., may, might, some, sometimes, usually, often, frequently, many, several, few, probably, likely).**
   Under various conditions, almost anything "might" happen so any answer could be right. These words can mean different things to different people.

9. **Item does not contain technical jargon, slang, or abbreviations or acronyms, unless they are uniformly and clearly understood.**
Use of these items often ends up testing something other than the content intended to be tested.

10. **Complex or K-Type items are not used.**
A complex or K-type item is one in which several words/statements are listed and the test-taker is asked to select from alternatives that include mixed combinations of the words/statements: (a) I and II are true, III is false, b) I and III are true, II is false c) I, II, and III are false, etc. These types of items are often confusing and test the test-takers' logic ability more so than content. As an example:

   A screening test for a disease has a sensitivity of 90% and a specificity of 95%, which means that of the people who had the test:

   1. 5% of those who have the disease were identified as positive

   2. 95% of those who do not have the disease were identified as negative

   3. 10% of those who were identified as negative have the disease

   4. 90% of those who were identified as positive do not have the disease

      a) 1 & 2 are true and 3 & 4 are false.

      b) 3 & 4 are true and 1 & 2 are false.

      c) 1 & 3 are true and 2 &4 are false.

      d) 2 &3 are true and 1 &4 are false.

11. **Item is written in the active, rather than passive voice (e.g., "The individual provided the instructions." (active) versus "The instructions were provided by the individual." (passive)**
The passive voice is not wrong, but the active voice is more concise and clear.

12. **Item consistently uses *either* general or specific terms, as appropriate, depending on subject matter being tested.**
Either can be used, but the key is to not mix general and specific terms within the same item (e.g., generic classifications of products versus brand names).

13. **Item repeats nouns, rather than using pronouns (e.g., it, they, he, she)**
If the noun to which the pronoun refers is perfectly clear, it is okay to use the pronoun. However, it is safest to just repeat the noun.

Stem Item Writing Guidelines

14. The stem should include only necessary information.
Extraneous information serves only to consume test-taker time and to confuse them, and therefore should be avoided. Sometimes it is tempting to add information to teach test takers, but the test is not an appropriate forum for teaching. Also, often this teaching material will clue an answer for a question elsewhere on the test.

15. The stem should be focused.
Each question should ask a single, complete, and focused question. A well-focused stem allows an examinee to determine what the question is asking for and formulate an answer before even looking at the alternatives. A test to see if the stem completely conveys the question is to ask if a knowledgeable examinee could answer the question without looking at the response choices. For example:

Poor: Which of the following statements is true about a patient with dysphagia?

Better: A 31-year-old woman has sudden onset of dysphagia and vomits all that she eats and drinks one year after undergoing a vertical-banded gastroplasty for morbid obesity. Which of the following is the most likely cause?

This first is vague. Its answer could be geographic, as in "based in the midwest" or be type of organization, as in "membership society, etc."   The second clearly asks what type of organization NIGP is. (Note that this question is for illustrative purposes, but still wouldn't be good to use since it assesses information not critical to practicing in the industry anyway!)

16. Stem is worded positively.
"Which of the following is not" and "all of the following except" ideally should not be used as negatively worded stems can inadvertently confuse test-takers. If they are used, the negative words should be <u>CLEARLY</u> identified. Never use a double negative (e.g., which of the following are *not* reasons to *not* hire a subcontractor?).

17. The stem should not contain new information.
All information that test takers need to answer a question should be listed before and separate from the stem itself. For example:

Poor: Which of the following is the most likely diagnosis for a 25-year-old man who has had weight loss, arthritis, and nocturnal diarrhea for the past year and for whom flexible sigmoidoscopy shows typical aphthous-type lesions?

Better: A 25-year-old man has had weight loss, arthritis, and nocturnal diarrhea for the past year. Flexible sigmoidoscopy shows typical aphthous-type lesions. Which of the following is the most likely diagnosis?

18. **In items calling for judgment, the authority or criterion that is the basis for the correct alternative (if there is one) is specified in the stem (e.g., "according to…").**

19. **The stem contains only necessary information.** For example:

Poor: The primary function of an organizational chart for a large non-profit primary care hospital is to show:

a) coordination between various departments

b) the chain of command and official channels.

c) responsibilities of the various departments.

d) characteristic tasks within the organization.

Better: The primary function of an organizational chart for a hospital is to show:

(same options)

The facts that the organization is large or non-profit or primary care have nothing to do with the issue being assessed, and should be, therefore, eliminated from the question.

20. **Stem does not include personal pronouns (e.g., you).**
Asking a question about" what should you do first" is more subjective than "what should a buyer do first". We are talking about what professionals should generally do in a described scenario, not what an individual (you) should specifically do (which conjures up their own organization's policies, politics, etc.).

Alternatives Item Writing Guidelines

21. There is only ONE correct or clearly best alternative.

22. Alternatives should be parallel in content.
A problem related to unfocused stems (mentioned earlier) is heterogeneous alternatives – those that relate to more than one topic. Unfocused stems usually lead to these wide-ranging alternatives. An item's alternatives should be homogeneous; they should be parallel in content. The following item violates this principle because the alternatives deal with four different topics.

> Which of the following statements about an ear of corn is true?
> *a)   It is considered a good source of calcium.
> b)   It grows only in Iowa.
> c)   It may be yellow or white.
> d)   It equals one vegetable exchange.

The stem does not alert the examinee about what is being asked—it could not, because the question is not focused. What is actually being asked is "What do you know about corn?" Many items phrased in the "Which of the following statements about _____ is true" or "_____ is:" format are not sufficiently focused, and the alternatives cover many topics. A better question would be:

> Corn is regarded as a good source of:
> a). iron.
> b) thiamin.
> *c) calcium.
> d) carbohydrate.

23. Alternatives should be believable.
In multiple-choice questions, the three incorrect alternatives should be believable, but also clearly wrong. Write distractors that are based on common misconceptions or miscalculations of less qualified or entry-level professionals.

24. "None of the above" and "All of the above" should not be used as alternatives, or should be used sparingly.
A test-taker need only know that one item is correct to select "None of the above", need only know that two of the alternatives are true to select "all of the above," and need only know that one alternative is wrong to eliminate "all of the above" as the correct answer.

25. Alternatives all use the same units or scale (e.g., metric).

26. The alternatives do not include any key words from the stem.
Repeating key words in alternatives can clue test-takers to the correct answer, or trick them into thinking a distractor is the correct answer.

27. Alternatives should be parallel in phrasing.

All alternatives should be written in the same format – all starting with a verb (same tense), noun, etc. An error that is common, but a definite give-away to test-takers is an incomplete statement format item with alternatives that do not conform grammatically to the stem.

For example:

Poor:

Alternatives should:

a) conform to the stem

b) *contains* all-inclusive words

c) be of differing lengths

Better:

Alternatives should:

a) conform to the stem

b) *contain* all-inclusive words

c) be of differing lengths

Poor:

The speaker's primary responsibility is to:

a) his/her audience's sense of humor

b) expense everything under the sun to the association

c) met learner needs

d) the hiring organization's contact

Better:

The speaker's primary responsibility is to:

a) make attendees laugh

b) meet contract obligations

c) fulfill learner needs

d) comply with staff requests

28. Alternatives should be mutually exclusive.

Items should contain alternatives that are mutually exclusive—that is, one alternative should not be the equivalent of, or contained within, any other alternative. Mutual exclusion is necessary so that only one alternative can be keyed as the correct answer. Although the other alternatives must be plausible, they must be clearly incorrect. For example:

Poor: 10-15
         15-17
         17-20
         20-22

Better: 10-14
           15-19
           20-24
           25-29

Poor: Under 100
         Under 200
         Under 300
         Under 400

Better: Under 100
           101-200
           201-300
           301-400

Poor: Increases buying	**Better:** Increases buying
Affects buying	Decreases buying

29. Alternatives should be presented in a logical order.

The alternative responses should be arranged in a logical order, if one exists (numerical, chronological, etc.). For example:

Poor:	10-14	**Better:**	10-14
	20-24		15-19
	15-19		20-24
	25-29		25-29

30. Alternatives should not clue the test taker.

Alternatives should not clue test-takers to the correct answer.

- Alternatives that contain all-inclusive words such as "all," "always," or "never" or "none" are usually incorrect so these words should be avoided when possible.

- Since an alternative that is much longer or more detailed than other ones is usually correct, try to keep the alternatives similar in length and detail.

- When two alternatives are opposites, usually one is correct. Don't use opposites, then, unless you use two pairs.