

Compact Test Item Writing Guide

Author: Tao Le, MD, MHS

Date last revised: September 9, 2009

Acknowledgements

Material in this guide was adapted with permission by MediQ Learning, LLC, from the USMLERx Item Writing Guide, 3rd Edition Revised, Copyright © 2007. Material in this guide was also adapted with permission by the National Board of Medical Examiners® (NBME®) from Constructing Written Test Questions for the Basic and Clinical Sciences, 3rd Edition Revised, Copyright © 2002 by the NBME. We also gratefully acknowledge the contributions of Kathy Harrington, PhD, MPH; Joan Mangan, PhD, MST, and Rachel Makleff, PhD, of the American Thoracic Society.

Audience

Presenters of Continuing Medical Education (CME) activities who wish to evaluate the competency of their participants with respect to their material prior to and/or after presenting an educational activity.

Purpose

This guide is intended to provide instructions for developing case vignette questions for evaluative purposes. Case vignettes allow for presentation of clinical and non-clinical scenarios with questions designed to test the competency of the participant specific to the material presented in the CME activity. Vignettes should be constructed to evaluate the learning objectives of the educational session to test competency post session (post-test only) or to test the effect of the educational activity in terms of improving competence in a specific area (pre and post-tests).

Question Construction

The question items are designed to assess competency using experimental and clinical vignettes. Each question should be constructed to test a learning objective of the educational activity. Learning objectives are the expected outcomes of the educational activity. In other words, they state what a participant should know, be able to apply or do as a result of participating in the learning activity. Each question is composed of a stem and a lead-in. The *stem* is the part of the question that sets the scenario/clinical vignette while the *lead-in* asks the question.

Example:

Question stem: A 35 year old machinist reports who is exposed to metal working fluids/coolants during a grinding process complains of a non-productive cough and shortness of breath. He has fever and chills. His chest x-ray demonstrates an interstitial infiltrate pattern and complete pulmonary function studies demonstrate a restrictive lung defect with no reversibility with bronchodilator. Pulse oximeter reads 95% on room air. He is a nonsmoker. He has experienced similar but less severe episodes over the past year.

Question lead in: What is the best initial treatment?

- A. Macrolide antibiotic
- B. Albuterol
- C. Fluticasone and salmeterol dry powder inhaler
- D. Prednisone*
- E. Cromolyn

Creating the Vignette Stem

Note that the clinical vignette stem may contain some of the following information:

1. Age, gender, race/ethnicity (e.g., 45 year-old Caucasian male)
2. Site of presentation (e.g., ED, clinic, hospital)
3. Presenting complaint (e.g., shortness of breath)
4. Duration (e.g., 2 weeks, three minutes)
5. Patient history (e.g., history of past myocardial infarction, smokes two packs each day)
6. Physical exam findings (e.g., wheezing)
7. Diagnostic tests (e.g., spirometry showed an FVC of 75%)
8. Initial treatment (e.g., albuterol nebulizer treatment, IM steroids)
9. Subsequent clinical course (e.g., patient continued to have marked shortness of breath)

An appropriately constructed question should have most of the content in the stem, with relatively brief lead-ins and answer choices. The stem should include all relevant facts; no additional data should be provided in the response options. Additionally, there should only be one lead-in per stem. Each question should be independent of the question following and preceding it. The ability to answer a question should not depend on how the participant answered the previous question.

Non-Clinical Vignette Questions

There are situations where non-clinical vignette questions are appropriate. These questions typically ask the learner to use their understanding of basic or clinical science, experimental design, and scientific tools of inquiry to predict or explain the results.

Example of a Non-Clinical Vignette Question:

A 24 year-old medical student is asked to breathe into a spirometer with a known volume and concentration of helium. She breathes in the spirometer until the helium concentration in the spirometer equals the concentration in her lungs. This technique allows the student to directly measure:

- A. Tidal volume
- B. Residual volume

- C. Functional residual volume**
- D. Vital capacity

Application of Knowledge

The question should ask for application of knowledge and not pure knowledge recall. Application of knowledge requires an examinee to reach a conclusion, make a prediction, or select a course of action. In addition, the question should be **directly relevant** to at least one stated learning objective of the educational activity. Application of knowledge questions allows the learner to express their information base as well as their ability to use that information. For example, look at the following rote memorization question and then the same type of question as an application of knowledge:

Examples:

Rote Memorization:

Acute intermittent porphyria is the result of a defect in the biosynthetic pathway for

- A. collagen
- B. corticosteroid
- C. fatty acid
- D. glucose
- E. heme**
- F. thyroxine (T4)

Application of Knowledge:

An otherwise healthy 33-year-old man has mild weakness and occasional episodes of steady, severe abdominal pain with some cramping but no diarrhea. One aunt and a cousin have had similar episodes. During an episode, his abdomen is distended, and bowel sounds are decreased. Neurologic examination shows mild weakness in the upper arms. *These findings suggest a defect in the biosynthetic pathway for....?*

- A. collagen
- B. corticosteroid
- C. fatty acid
- D. glucose
- E. heme**
- F. thyroxine (T4)

A question can integrate multiple basic and clinical science concepts, requiring multiple steps to solve. The following is an example of a multidisciplinary question:

A patient presents with malar rash, and signs of renal failure. She is given a common pharmacological treatment. What is a major side effect of this treatment?

This question has three logic steps:

1. Make the diagnosis (lupus erythematosus).
2. Know the common/preferred treatment (systemic corticosteroids).
3. Know the side effects of the common/preferred treatment (weight gain, increased appetite, insomnia, altered personality).

Creating the Question Lead-In

Now that you have created a stem, you must think about what type of *lead-in* you want to ask. The lead-in should be a short, concise, clear question that the examinee is able to answer without looking at the answer choices. Please do NOT write any lead-ins with the phrases: “Each of the following statements is correct EXCEPT?” or “Which of the following statements is correct/true/false?” Some sample lead-in categories include:

1. History and Physical

Examples:

Which part of the physical exam will help most in making the correct diagnosis?
Which of the following is the most important question to ask at this point in time?

2. Diagnostic Studies

Example:

Which diagnostic study will be most helpful in making the correct diagnosis?

3. Prognosis

In this type of question you will ask the participant to pick from a number of possible prognoses for a certain presentation.

Example:

What is the most likely prognosis?

4. Mechanisms of Disease and Basic Concepts

These items should test participants’ knowledge of pathophysiology in its broadest sense, including etiology, pathogenesis, natural history, clinical course, associated findings, complications, severity of illness, and intended or unintended effects of therapeutic interventions. Mechanism questions should be framed in a clinical context.

Examples:

Which of the following is the most likely explanation for these findings?
Which of the following is the most likely pathogen?
Which of the following findings is most likely to be increased/decreased?
A biopsy is most likely to show which of the following?

5. Diagnosis

For this category, write items that require participants to interpret history, physical findings, and results of laboratory, imaging, and other studies in order to determine the most likely diagnosis (differential diagnosis) or the next step in diagnosis (diagnostic testing).

Examples:

Which of the following is the most likely diagnosis?
Which of the following is most likely to confirm the diagnosis?

6. Management

These items assess principles of chronic and acute care in inpatient and outpatient settings. Questions cover pharmacotherapy decisions, more general treatment decisions where options include a mix of pharmacotherapy and other treatments, and management decisions where options include a mix of treatments and other things such as diagnostic testing.

Examples:

Which of the following is the most appropriate initial or next step in patient care?
Which of the following is the most effective management?
Which of the following is the most appropriate pharmacotherapy?
Which of the following is the first priority in caring for this patient? (e.g., in emergency department)

7. Data Interpretation

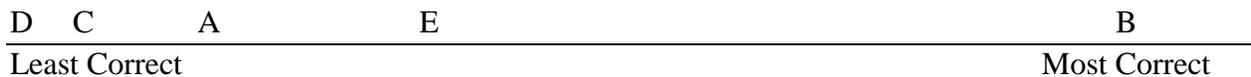
These items assess the participant’s ability to interpret laboratory or scientific data in the context of a scientific study

Examples:

Based on the data, what can be said about the relationship between Drug A and Drug B?
Based on the data, what conclusion can be drawn regarding Test A and Test B?

Answer Choice and Distracter Construction

All questions should be “one-best-answer” questions. Typically, there is one correct answer and four “distracters.” There will necessarily be four incorrect options, each of which is feasible, but less likely, choices. It should be possible to lay out the four incorrect answers and the correct answer on a continuum of “least correct” to “most correct.” Here, “B” is the correct answer:



The learner can then decide that all options are possible, with some more possible than others, but only one is the most likely. The options must differ on a single continuum (i.e., be answered without looking at the options) or else, must be 100% true or false. A good example of this is the following question:

Example:

Stem:

A 32-year-old man has a 4-day history of progressive weakness in his extremities. He has been healthy except for an upper respiratory tract infection 10 days ago. His temperature is 37.8 C (100 F), blood pressure is 130/80 mm Hg, pulse is 94/min, and respirations are 42/min and shallow. He has symmetric weakness of both sides of the face and the proximal and distal muscles of the extremities. Sensation is intact. No deep tendon reflexes can be elicited; the plantar responses are flexor.

Lead-in:

Which of the following is the most likely diagnosis?

Options:

- A. Acute disseminated encephalomyelitis
- B. Guillain-Barré syndrome
- C. Myasthenia gravis
- D. Poliomyelitis
- E. Polymyositis

Experts would all agree that the most likely diagnosis is B; they would also agree that the other diagnoses are somewhat likely (and would be included in a differential diagnosis), but less likely than B. The difficulty level of the questions is **directly** related to the distance between the best and the next best answer.

The following question is flawed:

Which of the following is true about pseudogout?

- A. It occurs frequently in women.*
- B. It is seldom associated with acute pain in a joint.*
- C. It may be associated with a finding of chondrocalcinosis.*
- D. It is clearly hereditary in most cases.*
- E. It responds well to treatment with allopurinol*

After reading the stem, the examinee has only the vaguest idea what the question is about. In an attempt to determine the “best” answer, the participants have to decide whether “it occurs frequently in women” is more or less true than “it is seldom associated with acute pain in a joint.” This is a comparison of apples and oranges. The options should be homogenous and mutually exclusive to one another. For example, it would be inappropriate to include individual drug treatments in the same option list as classes of drugs or drug dosages.

ITEM WRITING CHECKLIST

PART I -- GENERAL PRINCIPLES OF ITEM WRITING

1. Test application of knowledge using experimental and clinical vignettes.
2. Write items that have only one correct or clearly best answer.
3. Use items that measure important learning objectives. Focus items on key concepts and principles that are essential information (without access to references) for all participants to understand.
4. Write items that test material that is relevant to the specific setting in which that information is applicable.
5. Write items that require the examinee to use problem solving skills rather than fact regurgitation. Avoid items that *only* require recall of isolated facts.
6. Avoid esoteric or interesting topics that are not relevant to the learning objectives.
7. Write items that test only one concept.
8. All items need to be answered based on an understanding of presented material and basic underlying information the audience should have.
9. Avoid items containing ambiguous language or phrasing.
10. Follow standard rules of punctuation and grammar.
11. Use generic medication names only. Avoid brand names.
12. If items are based on opinion or authority such as a guideline, state whose opinion or what authority.
13. Avoid items based on statements taken verbatim from instructional materials.
14. Avoid items containing irrelevant clues.
15. Avoid items that have answers that are subject to change within a short period of time.
16. Avoid items that include “never” or “always.”

PART II -- MULTIPLE CHOICE ITEM WRITING

The Item Stem

1. Clearly present the information needed to answer the question in the item stem.
2. When writing the stem, do not leave a blank for completion in the beginning or middle of the stem.
3. Include only relevant material and necessary words in the stem.
4. Exclude grammatical cues in the item stem.
5. No two items should share the same stem.

The Item Lead-In

6. The lead-in should be a concise, clear question that the examinee is able to answer without looking at the answer choices. Please do NOT write any lead-ins with the phrases: “Each of the following statements is correct EXCEPT?” or “Which of the following statements is correct?”
7. Do not use negatively stated lead-ins such as “all of the following except...” or double negatives.
8. Do NOT write true/false items.
9. Vary the type of lead-in. You can ask about diagnoses, basic science principles, causes of a disease, functions, treatments, effects of a treatment, additional symptoms, biostatistics, behavioral science facts, etc.
10. Write integrated multidisciplinary items that require multistep logic, i.e., two or more steps of logic are needed to deduce the answer.

Answer Choices (The Options)

11. All items should be “one best answer” items.
12. Make all options homogeneous in content (e.g., all are diagnoses; all are next steps in patient care).
13. Write at least five options (one correct, four wrong answers) for every item.
14. Avoid correct options that are noticeably longer or shorter than the incorrect options.
15. Avoid stating the correct option in more detail than the incorrect options.
16. Avoid generalizing the correct option so that it has wider application than the incorrect options.
17. Avoid correct options that are one of two options that state the idea of fact in diametrically opposite fashion.
18. Avoid correct options that contain too familiar or stereotyped phraseology.
19. Make all options grammatically consistent with the stem.
20. Avoid incorrect options that contain language or technical terms that the examinee would not be expected to know.
21. Do not write items with the options all of the above or none of the above.
22. Make each distracter plausible and attractive to participants who have not mastered the material.
23. Make all options independent of each other.
24. Numerical answers should be stated consistently.
25. A subset of answer choices should not be exclusive.
26. Avoid specific determiners like “sometimes” and “never” in the options.
27. Use important-sounding words in the wrong options as well as in the correct option.
28. Avoid repeating key word(s) from the stem in the correct option.
29. Avoid incorrect options that are unreasonable options.