
AMERICAN BOARD OF DERMATOLOGY

QUESTION WRITING GUIDE

Image-Associated Questions

for

Part I Certifying Examination, In-Training Examination,
Maintenance of Certification Modules, and
Pediatric Dermatology Examination

Major portions reprinted with permission from “Constructing Written Test Questions for the Basic and Clinical Sciences” by Susan M. Case and David B. Swanson, Copyright©1996 by the National Board of Medical Examiners® (NBME®). The American Board of Dermatology gratefully acknowledges this permission.

CHAPTER I. INTRODUCTION

The preparation of good examination questions is much more difficult than most critics think. The natural tendency of new question writers is to prepare questions that are extremely difficult and which test small bits of obscure, rare information. Questions of this nature perform very poorly. Likewise, questions that are too easy, perform poorly because they do not discriminate between test takers. In reality, the best questions are those that 70-80% of the examinees will be able to answer correctly. This guide has been designed to help test committee members prepare better questions and incorporate both question theory and specific information on question writing. In preparing the guide, considerable material has been copied from National Board of Medical Examiners material, particularly the excellent booklet entitled, "Constructing Written Test Questions for the Basic and Clinical Sciences" by Susan Case and David B. Swanson.

The guide is designed to provide background material for the use of all test committees of the American Board of Dermatology (ABD). All ABD test committees use Type A questions in their examinations. Therefore, the chapters on Planning your Questions and Type A questions should be used by all test committee members. Everyone should also read the chapter on Technical Item Flaws. Your individual test committee chairman will inform you as to your need to review the chapters on Type B or R questions, as well as specific information on special needs such as photographs.

For image-associated questions the ABD currently uses two types of one-best-answer items. These are:

Type A - (5, or occasionally 4, options, single items or sets.) This is the type of question that is used for most questions on all of the examination.

Type R - (Extended-Matching items in sets of 2-20 items.) This type of question is used in the Part I examination, in the Dermatopathology examination and in the In-Training examination. *You will not be asked to write a Type R questions.*

Each of these question formats will be discussed separately below. The ABD does not use true-false-item formats that require that examinees select one or more options that are true.

CHAPTER II. PLANNING YOUR QUESTIONS

Before discussing specific question types, the question writer must understand the purpose of, and how to, structure questions to meet the goal of the examination.

Traditionally, test questions associated with images have been classified as requiring interpretation, or problem solving (comprehension, and reasoning) depending on the cognitive processes required to answer the question. "**Interpretation Questions**" require examinees to review the image, and reach some conclusion (e.g., a diagnosis). "**Problem-Solving Questions**" present an image and require examinees to take some action (e.g., the next step in patient management).

Writing application of knowledge questions is relatively straightforward in medicine. When you show an image of a patient and ask a question related to that patient, you are assessing application of knowledge. Application of knowledge questions are the types of questions that are appropriate for higher level examinations such as those of the ABD.

One approach to writing application of knowledge items is to use clinical vignettes. These vignettes might include an image with additional clinical and laboratory findings. The examinee may be asked to indicate the most likely diagnosis or to select the most appropriate next step in treatment or the laboratory study most likely to establish a diagnosis. The goal of this type of item is to have examinees synthesize information rather than merely recall isolated facts that can be looked up in a textbook.

Application of knowledge items can be structured to ask for:

1. A common factor (e.g., both x and y . . .)
2. That which fits or does not fit a classification
3. The effect or result of a change
4. A comparison, contrast, or analogy
5. A conclusion drawn from data (The most likely diagnosis is)
6. An illustration or an example
7. What is essential, basic, or necessary
8. The consequences of a possible change
9. An explanation of the changes
10. An application of a principle.
11. The implications of a decision
12. The most reasonable next step

CHAPTER III. TYPE A (ONE-BEST-ANSWER) QUESTIONS

Type A items consist of a stem (e.g., a clinical case presentation) and a lead-in question, followed by a series of five choices, typically one correct answer and four distractors. The task is to select the best response from among those offered. The following question describes a patient and asks the examinee to indicate the most likely diagnosis.

RULES FOR DEVELOPING A-TYPE IMAGE-ASSOCIATED ITEMS

1. ***Focus on an important concept, typically a common clinical concept or problem.*** Items should involve situations that a dermatologist would encounter in the context of practice. Avoid trivia or “tricky” or overly complex items.
2. ***Items should test application of knowledge, not recall of isolated facts.*** Item stems should be relatively short and options should be relatively short.
3. ***The item should be focused: the stem must pose a clear question, and it should be possible to arrive at an answer with the options covered.*** To determine whether you have written a focused item, cover the options and decide whether examinees who know the material could provide the single best answer based only on the stem. Rewrite the stem or options if they could not.

4. **All distractors (incorrect options) should be homogeneous; they should fall in the same category as the correct answer** (e.g., all diagnoses, tests, treatments, prognoses). All distractors should be both real and plausible, which implies that they be grammatically consistent, logically compatible, and of the same relative length as the answer.

Subject each item that you write to the rules outlined above. If an item follows all of these rules, it is probably well phrased and focused on an appropriate topic.

SUGGESTED STEMS FOR A-TYPE ITEMS associated with images

Following are some suggestions for developing items using images.

Basic Sciences; Mechanisms

Which of the following is the most likely explanation for these findings?
Which of the following is the most likely additional finding in this patient?
Laboratory evaluation is most likely to show
Which enzyme is most likely to be defective/deficient?
The most likely cause is occupational exposure to
This disorder is linked to an abnormality in which of the following genes?
The most likely cause is a deficiency in/a defect in the synthesis of
This electron micrograph depicts which of the following cells?

Diagnosis

Which of the following is the most likely diagnosis?
Which of the following is the most appropriate next step in establishing the diagnosis?
Which of the following findings will confirm the diagnosis?
This patient is at increased risk for development of
Which of the following patterns of immunoglobulin deposition is most likely?
Which of the following is the most likely causative organism?

Therapy

Which of the following is the most appropriate treatment?
Which of the following is the most appropriate initial step in management?
Which of the following is the mechanism of action?

DISTRACTORS

Rules for Developing Distractors

The incorrect options in each item are called distractors. In a well-constructed item, each distractor will be selected by some examinees. Therefore, all distractors should be plausible; none should stand out as being obviously incorrect. Common misconceptions, errors, and faulty reasoning provide a good source of plausible distractors.

Make sure that the distractors:

- a. are homogeneous in content (all are diagnoses, treatment options, laboratory values, etc.):
- b. are incorrect or definitely inferior to the correct answer;
- c. do not contain any clues to the correct answer;
- d. are real entities and would seem plausible and attractive to the uninformed;
- e. are similar to the correct answer in construction and length.

CHAPTER IV. TECHNICAL ITEM FLAWS

This section describes two types of technical item flaws: testwiseness and irrelevant difficulty. Flaws related to testwiseness make it easier for some students to answer the question correctly, based on their test-taking skills alone. These flaws commonly occur in items that are unfocused and do not satisfy the “cover-the-options” rule. Flaws related to irrelevant difficulty make the question difficult for reasons unrelated to the trait that is the focus of assessment.

The purpose of this section is to outline common flaws and to encourage you to eliminate these flaws from your questions to provide a level playing field for the testwise and not-so-testwise students. The probability of answering a question correctly should relate the examinee’s amount of expertise on the topic being assessed and should not relate to his or her expertise on test-taking strategies.

ISSUES RELATED TO TESTWISENESS

Absolute terms: terms such as “always” or “never” are used in options

In this item, Options A, B, and E contain terms that are less absolute than those in Options C and D. The testwise student will eliminate Options C and D as possibilities because they are less likely to be true than something stated less absolutely. Note that this flaw would not arise if the stem was focused and the options were short; it arises only when verbs are included in the options rather than in the stem.

In patients with advanced dementia, Alzheimer’s type, the memory defect

- A. *can be treated adequately with phosphatidylcholine (lecithin).*
- B. *could be a sequela of early parkinsonism.*
- C. *is never seen in patients with neurofibrillary tangles at autopsy.*
- D. *is never severe.*
- E. *possibly involves the cholinergic system.*

Long correct answer: correct answer is longer, more specific, or more complete than other options

In this item, Option C is longer than the other options; it is also the only double option. Item writers tend to pay more attention to the correct answer than to the distractors. Because you are teachers, you write long correct answers that include additional instructional material, parenthetical information, caveats, etc. Sometime this can be quite extreme: the correct answer is a paragraph in length and the distractors are single words.

Secondary gain is

- A. *synonymous with malingering.*
- B. *a frequent problem in obsessive-compulsive disorder.*
- C. *a complication of a variety of illnesses and tends to prolong many of them.*
- D. *never seen in organic brain damage.*

Word repeats: a word or phrase is included in the stem and in the correct answer

This item uses the word “unreal” in the stem and “derealization” is the correct answer. Sometimes, a word is repeated only in a metaphorical sense: a stem mentioning bone pain, with the correct answer beginning with osteo.

A 58-year-old man with a history of heavy alcohol use and previous psychiatric hospitalization is confused and agitated. He speaks of experiencing the world as unreal.

This symptom is called

- A. *derealization.*
- B. *depersonalization.*
- C. *derailment.*
- D. *focal memory deficit.*
- E. *signal anxiety.*

Frequency terms in the options are vague (e.g., rarely, usually)

Research has shown that vague frequency terms are not consistently defined, even by experts.

- Severe obesity in early adolescence*
- A. *usually responds dramatically to dietary regimens.*
 - B. *often is related to endocrine disorders.*
 - C. *has a 75% chance of clearing spontaneously.*
 - D. *shows a poor prognosis.*
 - E. *Usually responds to pharmacotherapy and intensive psychotherapy.*

Language in the options is not parallel; options are in a nonlogical order

This item illustrates a common flaw in which the options are long and the language makes it difficult and time-consuming to determine which is the most correct. Generally, this flaw can be corrected by careful editing. In this particular item, the lead-in can be changed to "For which of the following reasons can no conclusion be drawn from these data?" The options can then be edited (i.e., A. No follow-up was made of nonvaccinated children; B. The number of cases was too small; C. The trial involved only boys, and a new option can be written for D).

- In a vaccine trial, 200 2-year-old boys were given a vaccine against a certain disease and then monitored for five years for occurrence of the disease. Of this group, 85% never contracted the disease. Which of the following statements concerning these results is correct?*
- A. *No conclusion can be drawn, since no follow-up was made of nonvaccinated children.*
 - B. *The number of cases (i.e., 30 cases over five years) is too small for statistically meaningful conclusions.*
 - C. *No conclusions can be drawn because the trial involved only boys.*
 - D. *Vaccine efficacy (%) is calculated as $85-15/100$.*

None of the above is used as an option

The phrase "None of the above" is problematic in items where judgment is involved and where the options are not absolutely true or false. If the answer is intended to be one of the listed options, very knowledgeable students are faced with a dilemma, because they have to decide

- The diagnosis of a large ovarian cyst is most strongly suggested by an*
- A. *anterior dullness, lateral tympany.*
 - B. *decreased peristalsis.*
 - C. *fluid wave.*
 - D. *shifting dullness.*
 - E. *none of the above.*

between a very detailed perfect option and the one that you have developed as correct. They can generally construct an option that is more correct than the one you have intended to be correct. Use of "none of the above" essentially turns the item into a true/false item; each option has to be evaluated as more or less true than the universe of unlisted options.

SUMMARY OF TECHNICAL ITEM FLAWS

Issues Related to Testwiseness

- **Grammatical cues** - one or more distractors don't follow grammatically from the stem
- **Logical cues** - a subset of the options is collectively exhaustive
- **Absolute terms** - terms such as "always" or "never" are in some options
- **Long correct answer** - correct answer is longer, more specific, or more complete than other options
- **Word repeats** - a word or phrase is included in the stem and in the correct answer

Issues Related to Irrelevant Difficulty

- Options are long, complicated, or double
- Numeric data are not stated consistently
- Terms in the options are vague (e.g., "rarely," "usually")
- Language in the options is not parallel
- Options are in a nonlogical order
- "None of the above" is used as an option
- Stems are tricky or unnecessarily complicated
- The answer to an item is "hinged" to the answer of a related item

General Guidelines for Item Construction

- Make sure the item can be answered without looking at the options OR that the options are 100% true or false.
- Include as much of the item as possible in the stem; the stems should be long and the options short.
- Avoid superfluous information.
- Avoid "tricky" and overly complex items.
- Write options that are grammatically consistent and logically compatible with the stem; list them in logical or alphabetical order. Write distractors that are plausible and the same relative length as the answer.
- Avoid using absolutes such as *always*, *never*, and *all* in the options; also avoid using vague terms such as *usually* and *frequently*.
- Avoid negatively phrased items (e.g., those with *except* or *not* in the lead-in).

And most important of all: Focus on important concepts; don't waste time testing trivial facts.