
AMERICAN BOARD OF DERMATOLOGY

QUESTION WRITING GUIDE

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.

The ABD currently uses three 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 B - (4 or 5 option matching items in sets of 2-5 items.) This type of question is only in Part I, and the In-Training 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.

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 have been classified as requiring recall, interpretation, or problem solving (memory, comprehension, and reasoning) depending on the cognitive processes required to answer the question. Typical definitions refer to "Recall Questions" as those which assess knowledge of definitions or isolated facts. "Interpretation Questions" require examinees to review some information, often in tabular or graphic form, and reach some conclusion (e.g., a diagnosis). "Problem-Solving Questions" present a situation and require examinees to take some action (e.g., the next step in patient management). The difficulty with these classifications is that the cognitive processes required to answer a question are as dependent on the background of the examinee as they are on the question content. Experts in a content area may simply recall an answer with little or no conscious thought, whereas others may need to reason out the answer from basic principles. The cognitive processes involved in responding to a question are examinee-specific, making the taxonomic approach difficult to use.

An alternate approach divides items into two categories: application of knowledge or recall of an isolated fact. If a question requires an examinee to reach a conclusion, make a prediction, or select a course of action, it is classified as an *application of knowledge* question. If a question assesses only rote memory of an isolated fact (without requiring their application), it is classified as a recall question.

Another way of thinking about questions is to categorize them as either "top-down" or "bottom-up." Top-down questions are generally classified as knowledge or recall. They begin by citing a disease and then asking what patient findings are expected. These questions are structured similarly to most textbooks; the examinee could look up the disease and find the answer in a single paragraph. The flaw with "top-down" items is that they seem clinically backward. Patients rarely tell their physician what disease they have and then ask the physician what their signs and symptoms are.

Which of the following findings is most likely to be seen in postsurgical patients with pulmonary embolism?

In contrast, "bottom-up" questions are classified as application of knowledge. They are structured in a clinically more realistic manner, for example, by giving the findings and asking the examinee to indicate the underlying disease. Typically, examinees would need to be able to synthesize information from several pages of a textbook to answer these questions.

A 62-year-old man develops acute shortness of breath and pleuritic chest pain 4 hours after undergoing cholecystectomy. Which of the following is the most likely diagnosis?

Use of bottom-up questions with clinical vignettes as item stems has several benefits. First, the "face validity" of the examination is greatly enhanced by using questions that require examinees to "solve" clinical problems. Second, questions are more likely to focus on important information, rather than trivia. Third, these questions help to identify those examinees who have memorized a substantial body of factual information, but are unable to

use that information effectively in clinical situations. Bottom-up questions are generally more appropriate for higher level examinations.

Writing application of knowledge questions is relatively straightforward in medicine. When you describe 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 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.

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

Note that the incorrect options are not totally wrong. The options can be diagramed as follows:

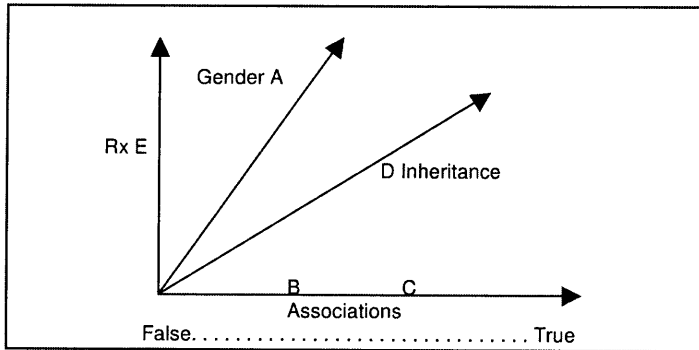


Even though the incorrect answers are not completely wrong, they are less correct than the “keyed answer.” The examinee is instructed to select the “most likely diagnosis”; experts would all agree that the most likely diagnosis is B; they would also agree that the other diagnoses are somewhat likely, but less likely than B. As long as the options can be laid out on a single continuum, in this case from “Most Likely Diagnosis” to “Least Likely Diagnosis,” options in one-best-answer questions do not have to be totally wrong.

This item is flawed. 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 examinees 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. In order to

- 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*

rank-order the relative correctness of options, they must differ on a single dimension or else all options must be absolutely 100% true or false.



The diagram of these options might look like this. The options are heterogeneous and deal with miscellaneous facts; they cannot be rank-ordered from least to most true along a single dimension. Although this question appears to assess knowledge of several different points, its inherent flaws preclude this. The question by itself is not clear; the item cannot be answered without looking at

the options.

In contrast to the options in the item on pseudogout, the options in the item on Guillain-Barré syndrome are all homogeneous (e.g., all diagnoses); knowledgeable examinees can rank-order the options along a single dimension.

Well-constructed one-best-answer questions satisfy the “cover-the-options” rule. They could be administered as write-in questions. The entire question is included in the stem.

It may be easier to write negative A-type questions, but this is discouraged by the ABD. The most problematic negative type-A questions are those that take the form: “Each of the following is correct EXCEPT” and “Which of the following statements is NOT correct?” These suffer from the same problem as true/false questions: if options cannot be rank-ordered on a single continuum, the examinees cannot determine either the “least” or the “most” correct answer.

RULES FOR DEVELOPING A-TYPE 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 long and options should be short. Vignettes provide a good basis for items that test application of knowledge. Each vignette should begin with a presenting problem and should be followed, where appropriate, by all or a subset of the following information: age and gender, history including duration of signs and symptoms, physical findings, results of diagnostic studies, treatment, subsequent problems. The stem should focus on important concepts rather than trivial facts and not be tricky or overly complex. The stem must include all relevant facts; no additional data should be provided in the options.

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) (see below). 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. In addition, options should be listed in logical or alphabetic order.
5. ***Avoid technical item flaws that provide benefit to test wise examinees or pose irrelevant difficulty.*** These flaws are discussed at length below.

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.

Do **NOT** write any items having the form “Which of the following statements concerning X is true?” or “Each of the following statements concerning X is correct EXCEPT.” These items are almost invariably unfocused, have heterogeneous options, and test isolated, often trivial facts.

SUGGESTED STEMS FOR A-TYPE ITEMS

Following are some suggestions for developing items using clinical vignettes. These items typically present a patient description that may include a combination of the following information: patient age, gender, physical signs and symptoms and their duration, results of laboratory studies and other diagnostic techniques. These descriptions are followed by a question or an incomplete statement such as one of the following:

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?*
- This agent acts at the receptor for*
- The most likely cause is occupational exposure to*
- What is the likelihood that other children will be affected?*
- The most likely cause is an embryologic defect in*
- This disorder is linked to an abnormality in which of the following genes?*
- This mass is composed primarily of*
- The zone of separation is between*
- The most likely cause is a deficiency in/a defect in the synthesis of*

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 findings differentiates disease X from disease Y?*
- 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 likely side effect of this treatment?*
- Which of the following is the most appropriate initial step in management?*
- Which of the following will decrease the risk for (complication X)?*
- Which of the following is the most likely outcome of (specific treatment)?*
- The interaction between these drugs is most likely to result in*
- 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.

Distractors directly affect the difficulty of an item. Consider the following question. Who was the primary author of the Declaration of Independence?

- A. Abraham Lincoln
- B. Thomas Jefferson
- C. Franklin Roosevelt
- D. King George III
- E. Catherine the Great

In this example, the options are quite divergent, and Thomas Jefferson is easily identified as the correct answer. Someone who knows relatively little about American history could answer this correctly.

Now consider the same questions with a different set of options:
Who was the primary author of the Declaration of Independence?

- A. George Washington
- B. Thomas Jefferson
- C. Alexander Hamilton
- D. Benjamin Franklin
- E. James Madison

In this example, the question becomes more difficult; the options are all plausible answers to someone who has limited knowledge.

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

Grammatical cues: one or more distractors don’t follow grammatically from the stem

Because an item writer tends to pay more attention to the correct answer than to the distractors, grammatical errors are more likely to occur in the distractors. In this example, testwise students would eliminate A and C as options because they do not follow grammatically or logically from the stem. Testwise students then have to choose only between B, D, and E.

A 60-year-old alcoholic derelict in status epilepticus is brought to the emergency department by the police. After ascertaining that the airway is open, the first step in management should be intravenous administration of

- A. *examination of cerebrospinal fluid.*
- B. *Glucose with vitamin B₁ (thiamine).*
- C. *CT scan of the head.*
- D. *phenytoin.*
- E. *diazepam.*

Logical cues: a subset of the options are collectively exhaustive

In this item, Options A, B, and C include all possibilities. The testwise student knows that A, B, or C must be correct, whereas the non-testwise student spends time considering D and E. Often, the item writers add D and E only because they want to list five options. In these situations, the item writer may not have paid much attention to the

Crime is

- A. *equally distributed among the social classes.*
- B. *over represented among the poor.*
- C. *over represented among the middle class and rich.*
- D. *primarily an indication of psychosexual maladjustment.*
- E. *Reaching a plateau of tolerability for the nation.*

merits of options D and E; sometimes, they are partially correct and confusing because they cannot be rank-ordered on the same dimension as Options A, B, and C. This flaw is

commonly seen in items with options such as “Increases,” “Decreases,” and “Remains the same.”

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.*

Convergence strategy: the correct answer includes the most elements in common with the other options

This item flaw is less obvious than the others, but it occurs frequently and is worth noting. The flaw is seen in several forms. The underlying premise is that the correct answer is the option that has the most in common with the other options; it is not likely to be an outlier. For example, in numeric options, the correct answer is more often the middle number than an extreme value. In double options, the correct answer is more likely to be the option that has the most common elements in common with the other distractors. For example, if the options are "Pencil and pen"; "Pencil and highlighter"; "Pencil and crayon"; "Pen and marker," the correct answer is likely to be "Pencil and pen" (i.e., by simple count, "Pencil" appeared 3 times in the options; "Pen" appeared twice; other elements each appeared only once).

- Local anesthetics are most effective in the*
- A. *anionic form, acting from inside the nerve membrane.*
 - B. *cationic form, acting from inside the nerve membrane.*
 - C. *cationic form, acting from outside the nerve membrane.*
 - D. *uncharged form, acting from inside the nerve membrane.*
 - E. *uncharged form, acting from outside the nerve membrane.*

While this might seem ridiculous, this flaw occurs because item writers start with the correct answer and write permutations of the correct answer as the distractors. The correct answer is, therefore, more likely to have elements in common with the rest of the options; the incorrect answers are more likely to be outliers as the item writer has difficulty generating viable distractors. In the example above, the testwise student would eliminate "anionic form" as unlikely; that student would also exclude "outside the nerve membrane." The student would then have to decide between Options B and D. Since four of the five options involve a charge, the testwise student would then pick Option B.

ISSUES RELATED TO IRRELEVANT DIFFICULTY

Options are long, complicated, or double

The item below illustrates a common flaw. The stem contains extraneous reading, but, more importantly, the options are very long and complicated. Trying to decide among these options requires a significant amount of reading because of the number of elements in each option. This can shift what is measured by an item from content knowledge to reading speed. Please note that this flaw relates only to options. There are many well-constructed test questions that include a long stem. Decisions about stem length should be made in accord with the purpose of the item. If the purpose of the item is to assess whether or not the student can interpret and synthesize information to determine, for example, the most likely diagnosis, then it is appropriate for the stem to include a fairly complete description of the patient.

Peer review committees in HMOs may move to take action against a physician's credentials to care for participants of the HMO. There is an associated requirement to assure that the physician receives due process in the course of these activities. Due process must include which of the following?

- A. Proper notice, a tribunal empowered to make the decision, a chance to confront witnesses against him/her, and chance to present evidence in defense.*
- B. Notice, an impartial forum, council, chance to hear and confront evidence against him/her.*
- C. Reasonable and timely notice, impartial panel empowered to make a decision, a chance to hear evidence against himself/herself and to confront witnesses, and the ability to present evidence in defense.*

Numeric data are not stated consistently

When numeric options are used, the options should be listed in numeric order and the options should be listed in a single format (i.e., as single terms or as ranges). Confusion occurs when formats are mixed and when the options are listed in an illogical order or in an inconsistent format.

In this example, Options A, B, and C are expressed as ranges, whereas Options D and E are specific percentages. All options should be expressed as ranges or as specific percentages; mixing them is ill-advised. In addition, the range for Option C includes Options D and E, which almost certainly rules out Options D and E as correct answers.

Following a second episode of salpingitis, what is the likelihood that a woman is infertile?

- A. Less than 20%*
- B. 20 to 30%*
- C. Greater than 50%*
- D. 90%*
- E. 75%*

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

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.

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.*

Stems are tricky or unnecessarily complicated

Sometimes, item writers can take a perfectly easy question and turn it into something so convoluted that only the most stalwart will even read it. This item is a sample of that kind of item.

Arrange the parents of the following children with Down's syndrome in order of highest to lowest risk of recurrence. Assume that the maternal age in all cases is 22 years and that a subsequent pregnancy occurs within 5 years. The karyotypes of the daughters are:

- I. *46, XX, -14, +T(14q21q) pat*
 - II. *46, XX, -14, +T(14q21q) de novo*
 - III. *46, XX, -14, +T(14q21q) mat*
 - IV. *46, XX, -21, +T(14q21q) pat*
 - V. *47, XX, -21, +T(21q21q) (parents not karyotyped)*
-
- A. *III, IV, I, V, II*
 - B. *IV, III, V, I, II*
 - C. *III, I, IV, V, II*
 - D. *IV, III, I, V, II*
 - E. *III, IV, I, II, V*

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
- **Convergence strategy** - the correct answer includes the most elements in common with the other options

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.

CHAPTER V. B-TYPE (MATCHING) ITEMS

Rules for Developing B-Type Items

1. If you are asked to write B-type questions, they should make up NO MORE THAN 20% of your assignment.
2. **Identify the theme for the set.** The themes may be topics such as chief complaint, diagnosis, appropriate therapy.
3. **Write a lead-in for the set.** The lead-in statement structures the task for the examinee and helps ensure that there is a single theme for the set. For example:
For each patient with headaches, select the most likely diagnosis.
4. **Prepare a list of options.** Four or preferably five brief homogeneous options should be used. For example:
 - A. Alprazolam
 - B. Flurazepam
 - C. Lorazepam
 - D. Oxazepam
 - E. Triazolam
5. **Write the items.** The items in each set should be similar in structure. If the items are patient vignettes, similar information should be included in each item. The patient's age, gender, and chief complaint should be listed in each option. In general, patients should be either all children or all adults because relevant options for one population are not necessarily relevant for the other.
6. **Review the items.** Check to assure that there is only a single "best" answer for each item. However, there is no need to have a perfect match. In fact, it is better to have more than one item match with a single option, thereby having one or more options that do not match items.

SAMPLE B-TYPE ITEMS

Example of a well-written B-type item:

For each patient, select the most appropriate drug class (A-E).

- A. Alpha-adrenergic agonists
- B. Alpha-adrenergic blockers
- C. Acetylcholinesterase inhibitors
- D. Beta-adrenergic agonists
- E. Beta-adrenergic blockers

1. A 40-year-old man has the sudden onset of severe headache, dizziness, and vomiting. His blood pressure is 260/130 mm Hg. He has encephalopathy and grade IV retinopathy.
2. A 55-year-old man has mild, chronic hypertension with tachycardia and other symptoms of mild congestive heart failure. He has a history of asthma and gout.

Example of a flawed B-type item:

For each of the following items, select the most closely associated option (A-E).

- A. Broad beans
- B. Dementia or toxic psychosis
- C. Electroconvulsive therapy
- D. Treatment for breast cancer
- E. Lithium carbonate

1. Digitalis
2. Succinylcholine
3. Tranylcypromine sulfate
4. Bromides
5. Fine tremor

Note: Both options and items contain a mixture of entities. The objective of the item set is not clear.

SUGGESTED SHELLS FOR B-TYPE ITEMS

Lead-in: For each patient with edema, select the antibody most specific for the symptoms described (A-E).

- Options:*
- A. Anti-centromere
 - B. Anti-neutrophil cytoplasmic
 - C.
 - D.
 - E.

Item: A 28-year-old woman has the recent onset of facial swelling, ankle edema, and joint stiffness. Urinalysis shows 3+ protein with a sediment containing 3 erythrocytes/hpf and 3 leukocytes/hpf. A 24-hour urine collection shows 8g of protein.

Lead-in: For each patient, select the most likely diagnosis (A-E).

- Options:*
- A. Anxiety disorder
 - B. Bipolar disorder, manic
 - C.
 - D.
 - E.

Item: A 45-year-old man with a history of alcoholism fidgets, is unable to sit still, and has difficulty concentrating.

Lead-in: For each patient with abnormal pigmentation, select the most likely diagnosis.

- Options:*
- A. Nevus anemicus
 - B. Nevus depigmentosus
 - C.
 - D.
 - E.

Item: A 19-year-old man has had a hypopigmented patch on the back since birth. Firm stroking of the lesion does not produce an axon flare.

CHAPTER VI. R-TYPE EXTENDED-MATCHING ITEMS

Extended Matching items are multiple-choice items organized into sets that use one list of options for all items in the set. A well-constructed Extended-Matching set includes four components: (1) a theme; (2) an option list; (3) a lead-in statement; and (4) at least two item stems, as illustrated below.

Theme: **Fatigue**

Options:

A.	Acute leukemia	H.	Hereditary spherocytosis
B.	Anemia of chronic disease I.		Hypothyroidism
C.	Congestive heart failure	J.	Iron deficiency
D.	Depression	K.	Lyme disease
E.	Epstein-Barr virus infectionL.		Microangiopathic hemolytic anemia
F.	Folate deficiency	M.	Miliary tuberculosis
G.	Glucose 6-phosphate dehydrogenase deficiency	N.	Vitamin B ₁₂ (cyanocobalamin) deficiency

Lead-in: For each patient with fatigue, select the most likely diagnosis.

Stems:

1. A 19-year-old woman has had fatigue, fever, and sore throat for the past week. She has a temperature of 38.3°C (101°F), cervical lymphadenopathy, and splenomegaly. Initial laboratory studies show a leukocyte count of 5000/mm³ (80% lymphocytes, with many lymphocytes exhibiting atypical features). Serum aspartate aminotransferase (AST, GOT) activity is 200 U/L. Serum bilirubin concentration and serum alkaline phosphatase activity are within normal limits.

Ans: E

2. A 15-year-old girl has a two-week history of fatigue and back pain. She has widespread bruising, pallor, and tenderness over the vertebrae and both femurs. Complete blood count shows hemoglobin concentration of 7.0 g/dL, leukocyte count of 2000/mm³, and platelet count of 15,000/mm³.

Ans: A

Extended-matching items are written differently than traditional one-best-answer items. Most often, the theme, lead-in, and options are written first; the items are written last. For example, if you want to write some questions related to the diagnosis of fatigue, you would begin by listing the diagnoses that might cause fatigue. You would then write a vignette for each of the options in the list. The example above includes vignettes for Epstein-Barr virus infection and for acute leukemia. Additional items might be written for some of the remaining diagnoses; for common, treatable diagnoses, more than one item might be prepared. The sample vignettes are moderate in length; shorter, more focused vignettes could also be used. Alternatively, examinees could be challenged to identify key diagnostic information, intermingled with incidental findings, by using longer vignettes.

WRITING EXTENDED-MATCHING ITEMS

The four components (theme, options, lead-in, and stems) are all essential for the construction of a good quality extended-matching set. Sets without lead-ins (or with nonspecific lead-ins, such as “Match each item with the best option”) should NOT be used, because they generally pose inconsistent or ambiguous tasks for examinees. The following set is flawed. The options are heterogeneous; there is no lead-in; the stems cannot be answered without reading the options. Rules for extended-matching items are completely analogous to those for one-best answer items.

Sample Extended-Matching Set - Flawed

- | | | | |
|----|-------------------------------------|----|---|
| A. | is motion sickness | I. | are completely controlled |
| B. | have no effects on people | J. | cause plant and eye damage |
| C. | indirectly increase CO ₂ | K. | are negligible |
| D. | cause death | L. | increase risk of skin cancer |
| E. | increased odor sensitivity | M. | cannot be controlled |
| F. | is a reduction in visibility | N. | excess acute respiratory illness among children |
| G. | esthetics, economics, health | O. | contrary to public opinion |
| H. | products of fossil fuel combustion | | |
-
1. Factors that people consider when evaluating air quality
 2. A principal effect of particulate matter in air
 3. The products of photochemical smog

After reading the stem in Item #1, examinees have only the vaguest idea what the question is about. In an attempt to determine the “best” answer, the examinees have to decide whether “is motion sickness” is more or less true than “have no effects on people.” The task is not do-able. Under these circumstances, unless an option is absolutely 100% true or false, it cannot be answered without looking at the options.

As with one-best-answer items, the stems should be long; the options should be short. There MUST be a lead-in that establishes the relationship between the items and the options. There should be NO verbs in the options. The “cover-the-options” rule is as relevant for extended-matching items as it is for one-best-answer items.

SAMPLE LEAD-INS AND TOPICS FOR OPTION LISTS

Patient vignettes provide an excellent structure for stems. Lead-ins generally begin with a phrase such as “For each of the following patients.” Often sets are organized around chief complaints or some other factor that allows a more specific introductory phrase such as “For each of the following patients with fatigue,” or “For each of the following patients with an enzyme deficiency.” The second part of the lead-in describes the task and the option set: “select the most likely diagnosis;” “select the protein that is most likely to be defective.”

The following are some additional sample lead-ins and some suggested topics for option lists.

- For each of the following patients, select the (e.g., nerve) that is most likely to be (abnormal/defective/deficient/nonfunctioning).
Options sets could include list of nerves; list of muscles; list of enzymes; list of hormones; list of proteins, list of types of cells; list of neurotransmitters; list of pathologic processes.
- For each of the following patients, select the (finding) that would be expected.
Options sets could include list of laboratory results; list of additional physical signs; autopsy results; results of microscopic examination of fluids, muscle or joint tissue; DNA analysis results; hormone levels.
- For each of the following patients, select the most likely (cause).
Options sets could include list of underlying mechanisms of the disease; medications that might cause side effects; list of drugs or drug classes; toxic agents, hemodynamic mechanisms.
- For each of the following patients, select the (e.g., drug) that should be administered.
Options sets could include list of drugs, vitamins, amino acids, enzymes, hormones.
- For each of the following patients with (chief complaint), select the most likely diagnosis.
Options sets could include list of diagnoses, most often organized around a chief complaint such as diseases that cause chest pain or diseases that cause fever.
- For each of the following patients, select the most appropriate next step in patient care.
Options sets could include list of pharmacologic therapies, list of laboratory studies, disposition alternatives, or the options could include a mixed set of treatments and additional studies to assess whether the student knows when sufficient data have been gathered.

MORE ON OPTIONS FOR R-SETS

Generally, anything that can be listed can form the basis for options in an R-set. Below are some topics that have been used as the basis for option lists.

Arteries	Connective Tissue Types
Nerves	Anatomic Structures
Muscles	Endocrine Structures
Amino Acids	Neurotransmitters
Peptides	Metabolic Defects
Hormones	Immune Disorders
Enzymes	Motor System Components
Cell Components	Cardiac Structures
Cell Types	Organelles
Blood Components	Congenital Anomalies
Molecules	Segments of the Spinal Cord
Karyotypes	Central Nervous System Components
Proteins	Secretory Products
Lipids	Extracellular Matrix Components
Pathogens/Bacteria/Fungi	Management Alternatives
Viruses	Drugs/Drug Classes
Cytokines	Pathologic Processes
Toxins	Pathophysiologic States
Vitamins/Minerals	Electrolyte Abnormalities
Diagnoses	Diagnostic Tests

The list of options should be single words or very short phrases. They must be homogeneous (all diagnoses, all management options, all anatomical sites, all vitamins, etc.) They can be labeled areas in a graph or in pictorial material. Options, especially those involving laboratory values, are often expressed in tabular form (see physiology example). Include all relevant options that are appropriate for the examinees; subtle distinctions and uncommon diagnoses may be inappropriate. For some topics, as few as three options might be appropriate; for others, a list of 26 (one for each letter in the alphabet) might be required.

WRITING THE ITEM STEMS

Each patient description should be similar in structure to the others in the set. For example, if race, ethnicity, or occupation is included in one item, it should be included in all items; if laboratory data are included for one item, include them in all items. It is advisable not to mix adults and pediatric cases in the same set - too often the age alone provides too much cueing and eliminates large numbers of options from consideration.

One advantage of the use of patient vignettes is that it helps to assure that the content assesses application of knowledge. These items should not resemble crossword puzzles, where both the options and the stems are single words or short phrases. Avoid restructuring those items you were faced with in junior high school where you had to draw a line from something in column A to the matching option in column B.

It is particularly important that the items be straightforward. There is no reason to make them tricky; the extended option list makes them difficult enough to allow you to distinguish the knowledgeable from the unknowledgeable student without resorting to trickery. As with well-constructed A-type questions, the "cover-the-options" rule is paramount. Knowledgeable students should be able to generate an answer to the question and then find that answer in the alphabetical list of options.

An item should be prepared for most of the options; for common or important options, more than one item can be written. In constructing an examination assessing general competence, to avoid overemphasizing a topic, all of the options, but only two or three of the items, would be used; the remaining items are retained for subsequent exams. On the other hand, if you want to assess knowledge in greater depth on a smaller number of topics, 10 to 20 items can be included for each set, with a subscore calculated for each topic.

In reviewing the items, check to make sure that there is only a single "best" answer for each question. Also make sure that there are at least four reasonable distractors for each item. As a final check, it is recommended that you ask a colleague to review the items (without the correct answer indicated). If the colleague has difficulty determining the correct answer, modify the option list or the item to eliminate the ambiguity.

SAMPLE GOOD AND BAD ITEM STEMS USING THE SAME OPTION LIST

The following is a good microbiology set. The options are a homogeneous list of pathogens; including both viruses and bacteria makes sense. There is a lead-in that presents a clear task for the examinee. There are three item stems that require students to apply their basic science knowledge or microbiology to arrive at the most likely cause of each patient's illness.

- | | | | |
|----|------------------------------------|----|--|
| A. | <i>Adenovirus</i> | L. | <i>Haemophilus influenzae</i> |
| B. | <i>Aspergillus fumigatus</i> | M. | <i>Histoplasma capsulatum</i> |
| C. | <i>Bacillus anthracis</i> | N. | <i>Mycobacterium tuberculosis</i> |
| D. | <i>Candida albicans</i> | O. | <i>Mycoplasma pneumoniae</i> |
| E. | <i>Chlamydia psitta</i> | P. | <i>Neisseria gonorrhoeae</i> |
| F. | <i>Coccidioides immitis</i> | Q. | <i>Neisseria meningitidis</i> |
| G. | Coronavirus | R. | <i>Pneumocystis carinii</i> |
| H. | <i>Corynebacterium diphtheriae</i> | S. | Rhinovirus |
| I. | <i>Coxiella burnetii</i> | T. | <i>Streptococcus pneumoniae</i> |
| J. | Coxsackievirus | U. | <i>Streptococcus pyogenes</i>
(group A) |
| K. | Epstein-Barr virus | | |

For each patient, select the pathogen most likely to have caused his/her illness.

1. A 7-year-old girl has a high fever and a sore throat. There is pharyngeal redness, a swollen right tonsil with creamy exudate, and painful right submandibular lymphadenopathy. Throat culture on blood agar yields numerous small β -hemolytic colonies that are inhibited by bacitracin.

Ans: U

2. For the past week, an 18-year-old man has had fever, sore throat, and malaise with bilaterally enlarged tonsils, tonsillar exudate, diffuse cervical lymphadenopathy, and splenomegaly. There is lymphocytosis with atypical lymphocytes. The patient tests positive for heterophil antibodies.

Ans: K

The following stem, developed for the same set, assesses recall of isolated facts rather than application of knowledge. It looks more like a crossword puzzle question than a question for a medical school examination.

3. An encapsulated gram-positive organism that usually grows in pairs or short chains.

OVERVIEW OF THE STEPS FOR WRITING EXTENDED-MATCHING ITEMS

1. **Identify the theme for the set.** The theme can be a chief complaint (e.g., chest pain, fatigue), a disposition situation (e.g., admission/discharge from the emergency department), a drug class (e.g., antihypertensive agents, antibiotics).
2. **Write the lead-in for the set** (e.g., *For each patient described below, select the most likely diagnosis*). The lead-in indicates the relationship between the stems and options, clarifying the question posed for examinees. It is an essential component of an Extended-Matching set.
3. **Prepare the list of options.** The list of options should be single words or very short phrases. List the options in alphabetical order unless there is a logical order.
4. **Write the items.** The items within a set should be similar in structure. Most often, patient vignettes are appropriate.
5. **Review the items.** Check to make sure that there is only a single “best” answer for each question. Also make sure that there are at least four reasonable distractors for each item. As a final check, it is recommended that you ask a colleague to review the items (without the correct answer indicated). If the colleague has difficulty determining the correct answer, modify the option list or the item to eliminate the ambiguity.

Additional information in writing Extended-Matching items can be found in:

Case SM, Swanson DB. Extended-matching items: a practical alternative to free-response questions. *Teaching and Learning in Medicine*. 1993;5(2):107-115.

Case SM, Swanson DB, Woolliscroft JO. Assessment of diagnostic pattern recognition skills in medicine clerkships using a written test. In: Harden R, Hart I, Mulholland H. eds., *Approaches to Assessment of Clinical Competence*. Norwich, England: Page Brothers; 1992: 452-458.

Sample Diagnosis Set

- | | | | |
|----|-------------------------------|----|---------------------------|
| A. | Ankylosing spondylitis | E. | Osteoporosis |
| B. | Intervertebral disc infection | F. | Spinal stenosis |
| C. | Multiple myeloma | G. | Spondylolysis |
| D. | Myofascial pain | H. | Tuberculosis of the spine |

For each patient with back pain, select the most likely diagnosis.

1. A 26-year-old man has insidious onset of low back pain and early morning stiffness. The pain alternates from side to side and occasionally radiates into the buttocks and back of the thighs, but not below the knees. The patient has acute anterior uveitis, diffuse low back and sacroiliac tenderness, and restricted range of motion of the hips. His erythrocyte sedimentation rate is 40 mm/h; latex fixation test is negative; and mild hypoproliferative anemia is present.

Ans: A

2. Twelve hours after being struck from the rear while driving her car, a 28-year-old woman has vague cervical and lumbar pain associated with headache and restricted cervical mobility. She is now very anxious. Rope-like bands of muscle are present in the lumbar area and over the left buttock; the bands are painful.

Ans: D

Additional items would cover some of the remaining diagnoses. The sample vignettes are of average length; shorter, more focused vignettes may also be used. Alternatively, examinees could be challenged to identify key diagnostic information by using longer vignettes.

Sample Diagnosis Set

- | | | | |
|----|-------------------------------|----|------------------------------|
| A. | Abdominal aneurysm | K. | Kidney stone |
| B. | Appendicitis | L. | Mesenteric adenitis |
| C. | Bowel obstruction | M. | Mesenteric artery thrombosis |
| D. | Cholecystitis | N. | Ovarian cyst -- ruptured |
| E. | Colon cancer | O. | Pancreatitis |
| F. | Constipation | P. | Pelvic inflammatory disease |
| G. | Diverticulitis | Q. | Peptic ulcer disease |
| H. | Endometriosis | R. | Perforated peptic ulcer |
| I. | Ectopic pregnancy -- ruptured | S. | Pyelonephritis |
| J. | Hernia | T. | Torsion |

For each patient with abdominal pain, select the most likely diagnosis.

1. A 25-year-old woman has sudden onset of persistent right lower abdominal pain that is increasing in severity. She has nausea without vomiting. She had a normal bowel movement just before onset of pain. Examination shows exquisite deep tenderness to palpation in right lower abdomen with guarding but no rebound; bowel sounds are present. Pelvic examination shows a 7-cm, exquisitely tender right sided mass. Hematocrit is 32%. Leukocyte count is $18,000/\text{mm}^3$. Serum amylase activity is within normal limits. Test of the stool for occult blood is negative.

Ans: B

2. An 84-year-old man in a nursing home has increasing left lower pain recurring every 3-4 hours over the past 3 days. He has no nausea or vomiting; the last bowel movement was not recorded. Examination shows a soft abdomen with a palpable, slightly tender, lower left abdominal mass. Hematocrit is 40%. Leukocyte count is $10,000/\text{mm}^3$. Serum amylase activity is within normal limits. Test of the stool for occult blood is negative.

Ans: E

Sample Management Set: Diagnostic Testing

- | | | | |
|----|------------------------------------|----|------------------------------|
| A. | Test of the stool for occult blood | H. | Exercise tolerance test |
| B. | Fasting serum glucose level | I. | Digital prostate examination |
| C. | Hemoglobin level | J. | ECG |
| D. | Prostate-specific antigen level | K. | Spirometry |
| E. | Serum cholesterol level | L. | X-ray film of the chest |
| F. | Serum iron level | M. | Sigmoidoscopy |
| G. | Thyroid function tests | | |

For each patient who comes to the physician for a health maintenance examination, select the most appropriate diagnostic study.

1. A 22-year-old man who weighs 89kg (196 lb) and is 175 cm (69 in) tall has smoked one pack of cigarettes daily for 8 years; he does not exercise. His last examination was 5 years ago. His father had a myocardial infarction at the age of 48 years. Physical examination shows no abnormalities.

Ans: E

2. A 28-year-old woman who weighs 70 kg (154 lb) and is 173 cm (68 in) tall has smoked one pack of cigarettes daily for 12 years; she does not exercise. Her last examination was 5 years ago, though she had a Pap smear 9 months ago that showed normal results. Her father had a myocardial infarction at the age of 48 years. Her grandfather was diagnosed with colon cancer at the age of 62 years. Physical examination shows no abnormalities.

Ans: E