

# Constructing Written Test Questions

Printed copies of this publication are not available from the National Board of Medical Examiners (NBME). Additional copies can be obtained by downloading the manual from the NBME's web site. Permission to copy and distribute this document is granted by the NBME provided that (1) the copyright and permission notices appear on all reproductions; (2) use of the document is for noncommercial, educational, and scientific purposes only; and (3) the document is not modified in any way. Any rights not expressly granted herein are reserved by the NBME.

Copyright © 1996, 1998 National Board of Medical Examiners® (NBME®).

Copyright © 2001, 2002 National Board of Medical Examiners® (NBME®).

All right reserved. Printed in the United States of America.

---

# Constructing Written Test Questions For the Basic and Clinical Sciences

**Third Edition**  
*(Revised)*

## *Contributing Authors*

*Susan M. Case, PhD and David B. Swanson, PhD\**

*National Board of Medical Examiners  
3750 Market Street  
Philadelphia, PA 19104*

*\*Dr. Case was a Senior Evaluation Officer at the National Board of Medical Examiners when the first three editions of this manual were published; she is currently the Director of Testing at the National Conference of Bar Examiners.*

*Dr. Swanson is the Deputy Vice President of Professional Services at the National Board of Medical Examiners.*

# Table of Contents

	<b>Page</b>
<b>Section I</b>	
<b>Issues Regarding Format and Structure of Test Questions</b> .....	7
<i>Chapter 1. Introduction</i> .....	9
Assessment: An Important Component of Instruction .....	9
Issues of Sampling .....	10
Importance of Psychometric Considerations .....	11
<i>Chapter 2. Multiple-Choice-Item Formats</i> .....	13
True/False vs One-Best-Answer Questions .....	13
The True/False Family .....	14
The One-Best-Answer Family .....	16
The Bottom Line on Item Formats .....	18
<i>Chapter 3. Technical Item Flaws</i> .....	19
Issues Related to Testwiseness .....	19
Issues Related to Irrelevant Difficulty .....	22
Summary of Technical Item Flaws .....	26
Use of Imprecise Terms in Examination Questions .....	27
<b>Section II</b>	
<b>Writing One-Best-Answer Questions for the Basic and Clinical Sciences</b> .....	31
The Basic Rules for One-Best-Answer Items .....	33
<i>Chapter 4. Item Content: Testing Application of Basic Science Knowledge</i> .....	35
Item Content for the Basic Sciences .....	35
Item Templates .....	38
Additional Templates .....	39
Types of Questions and Sample Lead-ins and Option Lists .....	40
Writing the Options: Altering Item Difficulty .....	41
Item Shape .....	37 451.9508 Tm(Sec

Testing Recall of Isolated Facts or Application of Knowledge .....	53
Writing One-Best-Answer Items .....	56
Fine Points on Item Stems .....	57
Verbosity, Window Dressing, and Red Herrings:	
Do They Make a Better Test Item? .....	58
Writing Items Related to Physician Tasks .....	61
Writing Items on Difficult Topics .....	66
<b>Section III</b>	
<b>Extended-Matching Items</b> .....	69
<i>Chapter 6. Extended-Matching (R-Type) Items</i> .....	71
Avoiding Flaws When You Write Extended-Matching Items for Your Own Examination .....	72
Sample Lead-ins and Topics for Option Lists .....	74
More on Options for R-Sets .....	75
Writing the Item Stems .....	76
Sample Good and Bad Item Stems Using the Same Option List .....	77
Overview of the Steps for Writing Extended-Matching Items .....	81
Sample Extended-Matching Sets .....	82
Steps for Organizing a Group to Write Clinical R-sets .....	90
Form for Writing R-Sets .....	93
Sample SPSSX Code to Score Multiple-Choice Tests Including Extended-Matching Items .....	94
Comparison of Items in Five-Option and Extended-Matching Format .....	96
A's to R's and Back Again .....	97
<i>Chapter 7. Pick N Items: An Extension of the Extended-Matching Format</i> .....	99
<b>Section IV</b>	
<b>Additional Issues</b> .....	105
<i>Chapter 8. Interpretation of Item Analysis Results</i> .....	107

---

## Preface to Third Edition

3

This manual was written to help faculty members improve the quality of the multiple-choice questions written for their examinations. The manual provides an overview of item formats, concentrating on the traditional one-best-answer and matching formats. It reviews issues related to technical item flaws and issues related to item content. The manual also provides basic information to help faculty review statistical indices of item quality after test administration. An overview of standard-setting techniques is also provided. Issues related to exam blueprinting are not addressed in any detail. We have focused almost exclusively on the item level, leaving exam level planning for another manuscript.

We anticipate that this manual will be useful primarily by faculty who are teaching medical students in basic science courses and clinical clerkships. The examples focus on undergraduate medical education, though the general approach to item writing may be useful for assessing examinees at other levels.

This manual reflects lessons that we have learned in developing items and tests over the past 20 years. During this period, we have reviewed (quite literally) tens of thousands of multiple-choice questions and have conducted item-writing workshops for thousands of item writers preparing USMLE, NBME, and specialty board examinations as well as faculty at more than 60 medical schools developing test questions for their own examinations. Each workshop attendee has helped us to frame our thoughts regarding how to write better quality test questions, and, over the years, we have become better able (we believe) to articulate the why's and wherefore's. We hope this manual helps to communicate these thoughts.

Susan M. Case, PhD  
David B. Swanson, PhD

January 1998

---



---

# **Section I**

## **Issues Regarding Format and Structure of Test Questions**

3

This section reviews structural issues important for the construction of high-quality test questions. The following section will review issues related to item content.





# Chapter 1

## Introduction

3

### **Assessment: An Important Component of Instruction**

Assessment is a critical component of instruction; properly used, it can aid in accomplishing key curricular goals. The impact of decisions regarding how and when to evaluate the knowledge and performance of your students cannot be overestimated.

A primary purpose of testing is to communicate what you view as important. Tests are a powerful motivator, and students will learn what they believe you value. Assessment also helps to fill instructional gaps by encouraging students to read broadly on their own and to participate broadly as educational opportunities are available. This outcome of testing is especially important in the clerkships, where the curriculum may vary from student to student, depending on factors such as the clinical setting and the random flow of patients. This outcome may also be important in some basic-science settings (eg, problem-based learning), where the educational experiences may vary from student to student.

Because tests have such a powerful influence on student learning, it is important to develop tests that will further your educational goals. Introduction of a hands-on clinical skills test drives students out of the library into the clinic, where they may seek help with their physical-exam skills; introduction of a test assessing only recall of isolated facts, on the other hand, drives them to “cram” course review books. This manual focuses on how to write high-quality, multiple-choice questions that

## Purposes of Testing

- Communicate to students what material is important
- Motivate students to study
- Identify areas of deficiency in need of remediation or further learning
- Determine final grades or make promotion decisions
- Identify areas where the course/curriculum is weak

## What Should Be Tested?

- Exam content should match course/clerkship objectives
- Important topics should be weighted more heavily than less important topics
- The testing time devoted to each topic should reflect the relative importance of the topic
- The sample of items should be representative of the instructional goals

## Issues of Sampling

The purpose of any assessment is to permit inferences to be drawn concerning the skills of examinees: inferences that extend beyond the particular problems (or, equivalently, cases or test questions) included in the exam to the larger domain from which the cases (or questions) are sampled. It's clear to all of us that assessment takes time. It's also clear that, if you increase time spent in one activity, you have to decrease time spent in other activities. Whether you're deciding on an overall plan for evaluation or you're deciding what to include on a single test, you're basically faced with a sampling problem. Performance on the sample provides a basis for estimating achievement in the broader domain that is actually of interest.

With multiple-choice questions (MCQs), you first need to decide what you want to include on the test. The amount of attention given to evaluating something should reflect its relative importance. You need to sample topics and also sample skills (eg, determining the diagnosis, deciding on the next step in management); you cannot ask everything. Performance on the sample provides a basis for estimating achievement in the broader domain that is actually of interest. The nature of the sample determines the extent to which the estimate of true ability is reproducible (reliable, generalizable) and accurate (valid). If the sample is not representative of the broader domain of interest (eg, including only cardiovascular-related content in a

test of competence in general medical practice), exam results will be biased and will not provide a good basis for estimating achievement in the domain of interest. If the sample is too small, exam results may not be sufficiently precise (reproducible, reliable) to ensure that they reflect true proficiency.

With a multiple-choice test, there's almost always one grader (usually the computer) and a series of questions or sets of questions; sampling involves selecting a subset of questions to include on the test. With other evaluation methods (eg, oral exams based on patient cases, standardized patient exams, essay exams), the sampling is much more complicated. Any method that can't be scored mechanically requires sampling on a second dimension: the dimension of grader. In these exams, you are interested in performance across a range of cases and you want the grade to be independent of who the examiner is. You therefore need to sample across two dimensions: one for the questions or cases and one for the judges or raters. You need to sample across a range of cases, because performance on one case is not a very good predictor of performance on other cases. You also need to sample across different raters to minimize the effects of rater harshness or leniency, and other issues like halo that cause problems in the consistency of scoring across raters. With broad samples, peaks and valleys in performance and peaks and valleys in rater differences tend to average out.

Although this manual focuses on multiple-choice questions, we believe that it is generally appropriate to use a variety of testing methods. No one method is likely to assess all the skills of interest. It should also be noted that the method used for assessment does not directly affect test quality, nor does it determine the component of competence measured by the test.

## Importance of Psychometric Considerations

The extent to which the psychometric characteristics of an assessment method are important is determined by the purpose of the test and the decisions that will be made based on the results. For "high-stakes" tests (those used for promotion or graduation decisions, even course grades), test results must be reasonably reproducible (precise, reliable) and accurate (valid). For "low-stakes" tests, the psychometric characteristics are less important, and the primary consideration should be on directing student learning. As noted above, in order to generate a reproducible score, you need to sample content broadly (ie, typically, a dozen or more cases, 100 or more MCQs or short-answer items).

The following papers include more detail about assessment issues in general.

Swanson DB. (1987) A measurement framework for performance-based tests. In Hart I, Harden R (Eds.), *Further Developments in Assessing Clinical Competence*. Montreal: Can-Heal Publications, 13-4.

Swanson DB, Norcini JJ, Grosso L. (1987) Assessment of clinical competence: written and computer-based simulations. *Assessment and Evaluation in Higher Education*, 12(3), 220-246.

Friedman C, de Blik R, Greer D, Mennin S, Norman G, Sheps C, Swanson DB, Woodward C. (1990) Charting the winds of change: recommendations for evaluating innovative medical curricula. *Academic Medicine*, 65, 8-14.

Swanson DB, Case SM, van der Vleuten CP. (1991:1997/2nd edition) Strategies for student assessment. In Boud, D. & Feletti, G (Eds.), *The Challenge of Problem-Based Learning*. London, Kogan Page Ltd., 269-282.

Newble DI, Dauphinee D, Woolliscroft JO, MacDonald M, Mulholland H, Page G, Swanson DB, Thomson A, van der Vleuten CP. (1994) Guidelines for assessing clinical competence. *Teaching and Learning in Medicine*, 6:3, 213-220.

Swanson DB, Norman G, Linn R. (1995) Performance-based assessment: Lessons from the health professions. *Educational Researcher*, 24:5, pp5-

# Chapter 2

## Multiple-Choice-Item Formats

3

In order for a test question to be a good one, it must satisfy two basic criteria. First, the test question must address impor-

## **The True/False Family**

The true/false and one-best-answer families pose very different tasks for the examinee. True/false items require an examinee to select all the options that are “true.” For these items, the examinee must decide where to make the cut-off — to what extent must a response be “true” in order to be keyed as “true.” While this task requires additional judgement (beyond what is required in selecting the one best answer), this additional judgment may be unrelated to clinical expertise or knowledge. Too often, examinees have to guess what the item writer had in mind because the options are not either completely true or completely false.

While written in jest (by the second author), this true/false item illustrates a common problem — items for which the stem is unclear. Depending on your perspective, Options 1, 2, and 3 might be true; alternatively, 1, 2, and 3 might be false while 4 is true.

***The way to a man's heart is through his***

1. *aorta*
2. *pulmonary arteries*
3. *pulmonary veins*
4. *stomach*

In this true/false example, there are vague terms in the options that provide cues to the testwise examinee. For example, the term “may” in Options 1, 2, and 3 cues the testwise examinee that those options are true. Option 4 is harder to guess — what does “usually” mean? Research has shown that these vague frequency terms do not have a shared definition. Experts would not agree on whether the fourth option is true or false.

***In the clinical assessment of chronic pain,***

1. *the physician's personal attitude concerning pain may affect medical judgement*
2. *unpleasant emotions may be converted to complaints of bodily pain*
3. *pain may have a symbolic meaning*
4. *facial appearance or body posture is usually a clue to the severity of the pain*

The flaws in this item are more subtle. The difficulty is that the examinee has to make assumptions about the severity of the disease, the age of the patient, and whether or not the disease has been treated. Different assumptions lead to different answers, even among experts.

***In children, ventricular septal defects are associated with***

1. *systolic murmur*
2. *pulmonary hypertension*
3. *tetralogy of Fallot*
4. *cyanosis*

Note that in each sample flawed item, the stem is unclear, the options contain vague terms, or the options are partially correct. In each instance, a group of experts would have difficulty reaching a consensus on the correct answer.



Because examinees are required to select all the options that are “true,” true/false items must satisfy the following rules:

- Stems must be clear and unambiguous. Imprecise phrases such as *is associated with*; *is useful for*; *is important* and words that provide cueing such as *may* or *could be*; and vague terms such as *usually* or *frequently* should be avoided.
- Options must be absolutely true or false; no shades of gray are permissible; avoid phrases and words noted in the first item above.

## The One-Best-Answer Family

In contrast to true/false questions, one-best-answer (A-type) questions make explicit the number of options to be selected. A-type items are the most widely used multiple-choice-item format. They consist of a stem (eg, a clinical case presentation) and a lead-in question, followed by a series of choices, typically one correct answer and four distractors. The following question describes a situation (in this instance, a patient) and asks the examinee to indicate the most likely cause of the problem.

### **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





---

## Chapter 3

# Technical Item Flaws

3

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 to the examinee’s amount of expertise on the topic being assessed and should not relate to their 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 man is brought to the emergency department by the police, who found him lying unconscious on the sidewalk. 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 B<sub>1</sub> (thiamine)*
- C. *CT scan of the head*
- D. *phenytoin*
- E. *diazepam*



**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. Sometimes 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, eg, a stem mentioning bone pain, with the correct answer beginning with the prefix “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. *depersonalization*
- B. *derailment*
- C. *derealization*
- 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 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” (ie, by simple count, “Pencil” appeared 3 times in the options; “Pen” appeared twice; other elements each appeared only once). 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 this example, the testwise student would eliminate “anionic form” as unlikely because “anionic form” appears only once; that student would also exclude “outside the nerve membrane” because “outside” appears less frequently than “inside”. The student would then have to decide between Options B and D. Since three of the five options involve a charge, the testwise student would then pick Option B.

***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*

## Issues Related to Irrelevant Difficulty

### Options are long, complicated, or double

This item 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 situation.



### **Frequency terms in the options are vague (eg, rarely, usually)**

Research has shown that vague frequency terms are not consistently defined or interpreted, even by experts. A more complete discussion of this research is included on page 29.

### **Language in the options is not parallel; options are in an illogical 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 results?” The options can then be edited (ie, 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?*

### “None of the above” is used as an option

The phrase “None of the above” is problematic in items where judgement is involved and where the options are not absolutely true or false. If the correct response is intended to be one of the other listed options, knowledgeable students can be faced with a dilemma because they have to decide between a very detailed perfect option and the one that you have intended as correct. They can often construct an option that is more correct than the one you 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. It will often be possible to fix such items by replacing “none of the above”

by an option that means roughly the same thing but is more specific. For example, in an item asking an examinee to specify the most appropriate pharmacotherapy, replacing “none of the above” by “no drug should be given at this time” will eliminate the ambiguity of “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. The notation in I: through V: is complex; having to rank order Roman numerals after working through that notation is irrelevant and unnecessarily difficult.

**Which city is closest to New York City?**

- A. Boston
- B. Chicago
- C. Dallas
- D. Los Angeles
- E. none of the above

*If students select E, you don't know if they are thinking about Philadelphia or London.*

**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 (eg, "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 (eg, those with *except* or *not* in the lead-in). If you must use a negative stem, use only short (preferably single word) options.

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

## Use of Imprecise Terms in Examination Questions

While imprecise terms are used in our everyday speech and in our writing, these terms cause confusion when they are used in the text of examination items. In a study conducted at the NBME, 60 members of eight test committees who wrote questions for various medical specialty examinations reviewed a list of terms used in MCQs to express some concept related to frequency of occurrence and indicated the percentage of time that was reflected by each term.

Results (shown below) indicated that the terms do not have an operational definition that is commonly shared, even among the item writers themselves. The mean value plus or minus one standard deviation exceeded 50 percentage points for more than half of the phrases. For example, on average, the item writers believed the term *frequently* indicated 70% of the time; half believed it was between 45% and 75% of the time; actual responses ranged from 20% to 80%. Of particular note is that values for *frequently* overlapped with values for *rarely*.

The implication of these results for the construction of test questions varies by item format. Vague terms create far more severe problems in the various kinds of true/false items (K-, C- and X-type items) than in one-best-answer (A- and R-type) items. For example, imprecise terms cause major problems in true/false items such as this example:

Rewording the options by specifying exact numbers does not correct the problem. For example, the statement, “the incidence among women is 1:2000” would not be an appropriate modification of Option 1 in the example shown. The incidence is not exactly 1:2000, and because a band is not specified, examinees would define their own bands, narrowly or widely, presumably depending on personal response styles. In true/false items, the appropriate treatment of numeric options is either to generate a comparison (eg, the incidence is greater than that of osteoarthritis) or to specify a range (eg, the incidence is between 1:1000 and 1:2000).

The issue noted above with true/false items is not as problematic with well-constructed one-best-answer items (ie, those that pose a clear question and have homogeneous options). For example, the following question includes a vague term in the item stem, yet, because the task is to select the one-best answer, the question is relatively unambiguous.

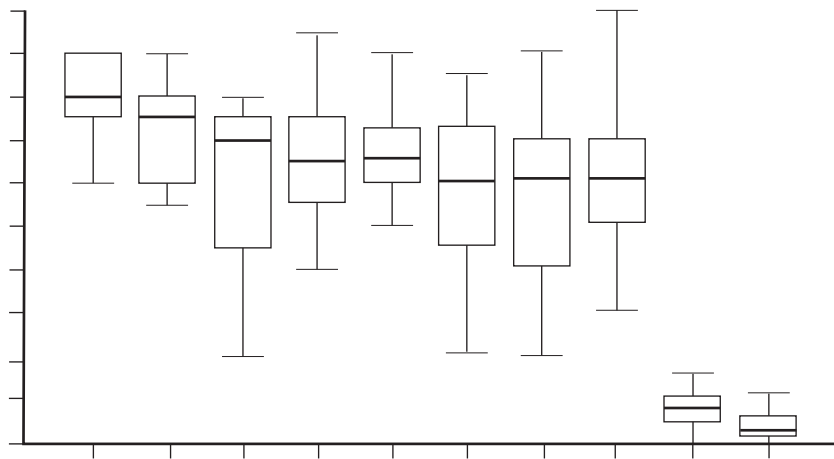
Which of the following laboratory values is usually increased in patients with pseudogout?

Problems do arise with one-best answer items that have vague terms in the options as in this example.

The only way to make such an item more ambiguous would be to use a fifth option “none of the above.”

***Patients with pseudogout have pain:***

- A. *frequently*
- B. *usually*
- C. *often*
- D. *commonly*



Box-plot showing distribution of responses for frequency terms. Results are based on responses from 60 members of eight item-writing committees. The horizontal line in each box indicates the median response; the boxes include the ranges for 50% of the responses. The vertical lines extend to the highest and lowest values indicated. For example, the median response for “frequently” indicated 70% of the time; half believed it was between 45% and 75% of the time; actual responses ranged from 20% to 80%, almost overlapping with “rarely.”

From: Case SM. (1994) The use of imprecise terms in examination questions: How frequent is frequently? *Academic Medicine*, 69(suppl):S4-S6.



## **Section II**

# **Writing One-Best-Answer Questions for the Basic and Clinical Sciences**

3

The previous chapters outlined technical issues related to the construc-





## The Basic Rules for One-Best-Answer Items

- ***Each item should focus on an important concept, typically a common or potentially catastrophic clinical problem.*** Don't waste testing time with questions assessing knowledge of trivial facts. Focus on problems that would be encountered in real life. Avoid trivial, "tricky," or overly complex questions.
- ***Each item should assess application of knowledge, not recall of an isolated fact.*** The item stems may be relatively long; the options should be short. Clinical vignettes provide a good basis for a question. For the clinical sciences, each should begin with the presenting problem of a patient, followed by the history (including duration of signs and symptoms), physical findings, results of diagnostic studies, initial treatment, subsequent findings, etc. Vignettes may include only a subset of this information, but the information should be provided in this specified order. For the basic sciences, patient vignettes may be very brief; "laboratory vignettes" are also appropriate.
- ***The stem of the item must pose a clear question, and it should be possible to arrive at an answer with the options covered.*** To determine if the question is focused, cover up the options and see if the question is clear and if the examinees can pose an answer based only on the stem. Rewrite the stem and/or options if they could not.
- ***All distractors (ie, incorrect options) should be homogeneous.*** They should fall into the same category as the correct answer (eg, all diagnoses, tests, treatments, prognoses, disposition alternatives). Rewrite any dissimilar distractors. Avoid using "double options" (eg, do W and X; do Y because of Z) unless the correct answer and all distractors are double options. Rewrite double options to focus on a single point. All distractors should be plausible, grammatically consistent, logically compatible, and of the same (relative) length as the correct answer. Order the options in logical order (eg, numeric), or in alphabetical order.
- ***Avoid technical item flaws that provide special benefit to testwise examinees or that pose irrelevant difficulty.***

Do **NOT** write any questions of the form "Which of the following statements is correct?" or "Each of the following statements is correct EXCEPT." These questions are unfocused and have heterogeneous options.

*Subject each question to the five "tests" implied by the above rules. If a question passes all five, it is probably well-phrased and focused on an appropriate topic.*

See also:

Swanson DB, Case SM. Assessment in basic science instruction: directions for practice and research. *Advances in Health Sciences Education: Theory and Practice*. 1997; 2:71-84.

**Chapter 4**  
**Item Content:**  
**Testing Application of Basic Science Knowledge**

3

The following pair of item stems illustrate the difference between a question assessing recall of an isolated fact and a question assessing application of knowledge.

**Basic Science Recall Item Stem:**

## ***Guidelines for Basic Science Item Content***

- Test application of knowledge using experimental and clinical vignettes
- Focus items on key concepts and principles that are essential information (without access to references) for all examinees to understand
- Test material that is relevant to learning in clinical clerkships, postgraduate medical education, and beyond
- Avoid items that *only* require recall of isolated facts
- Avoid esoteric or interesting topics that are not essential

These two items were written to assess the same topic. We recommend that questions be written like the second item, not the first one.

***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 ( $T_4$ )

***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 ( $T_4$ )

## Item Templates

The overall structure of an item can be depicted by an item template. You can typically generate many items using the same template. For example, the following template could be used to generate a series of questions related to gross anatomy:

*A (patient description) is unable to (functional disability). Which of the following is most likely to have been injured?*

This is a question that could be written using this template:

***A 65-year-old man has difficulty rising from a seated position and straightening his trunk, but he has no difficulty flexing his leg. Which of the following muscles is most likely to have been injured?***

- \*A. Gluteus maximus*
- B. Gluteus minimus*
- C. Hamstrings*
- D. Iliopsoas*
- E. Obturator internus*

## Additional Templates

A (*patient description*) has a (*type of injury and location*). Which of the following structures is most likely to be affected?

A (*patient description*) has (*history findings*) and is taking (*medications*). Which of the following medications is the most likely cause of his (*one history, PE or lab finding*)?

A (*patient description*) has (*abnormal findings*)



## Types of Questions

Guess my drug  
Guess my toxic exposure  
Guess my diet  
Guess my mood  
  
Predict physical findings  
Predict lab findings  
Predict sequelae  
  
Identify underlying cause/diagnosis  
Identify cause of drug responses  
Identify drug to administer

## Sample Lead-ins and Option Lists

Which of the following is (abnormal)?

*Options sets could include sites of lesions; list of nerves; list of muscles; list of enzymes; list of hormones; types of cells; list of neurotransmitters; list of toxins, molecules, vessels, spinal segments.*

Which of the following findings is most likely?

*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; serum levels.*

Which of the following is the most likely cause?

*Options sets could include list of underlying mechanisms of the disease; medications that might cause side effects; drugs or drug classes; toxic agents; hemodynamic mechanisms, viruses, metabolic defects.*

Which of the following should be administered?

*Options sets could include drugs, vitamins, amino acids, enzymes, hormones.*

Which of the following is defective/deficient/nonfunctioning?

*Options sets could include list of enzymes, feedback mechanisms, endocrine structures, dietary elements, vitamins.*

Given the pedigree, what is the likelihood that the next child (specify gender) will have the disease?

## Writing the Options: Altering Item Difficulty

The incorrect options in each question are called distractors. Each distractor should be selected by some examinees; therefore, each distractor should be plausible and none should stand out as being obviously incorrect. Common misconceptions and faulty reasoning provide a good source of plausible distractors. Distractors directly affect the difficulty of a question. Consider the question to the right.

***Who was the primary author of the Declaration of Independence?***

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

In the example above, 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 question with a different set of options.

In this example, the question becomes more difficult; the options are all plausible answers to someone who has limited knowledge. For some content areas, options like those in the first example might be appropriate; for others, those in the second example are more appropriate.

***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*

When writing your options, make sure that they are:

- Homogeneous in content (eg, all are diagnoses; all are next steps in patient care)
- Incorrect or inferior to the correct answer
- Plausible and attractive to the uninformed
- Similar to the correct answer in construction and length
- Grammatically consistent and logically compatible with the stem

## Item Shape

An appropriately shaped item includes as much of the item as possible in the stem; the stem should be relatively long and the options should be relatively short. The stem should include all relevant facts; no additional data should be provided in the options.

### Appropriately Shaped Item:

#### Long Stem

- A.
- B.
- C. Short Options
- D.
- E.

### Poorly Shaped Item:

#### Short Stem

- A.
- B.
- C. Long Options
- D.
- E.

## Problem-Based Learning and Use of Case Clusters

An increasing number of medical schools have adopted problem-based learning (PBL) as an instructional strategy for portions of the basic science curriculum. Although each school's approach to PBL is somewhat unique, all involve the use of written patient cases (problems) in basic science instruction. Problems are designed to stimulate learning of material from traditional basic science disciplines (eg, anatomy, physiology, biochemistry) from a clinical perspective, and application of basic science principles to clinical situations is stressed. Material is typically covered through independent study and discussed in small groups with a faculty tutor.

PBL courses and curricula typically emphasize the learning process, learning how to learn, responsibility of students for their own learning, and preparation for lifelong learning. However, there are important variations among programs that have implications for assessment. The Open Discovery approach emphasizes the learning process: students have responsibility for determining what to learn, as well as when and how to learn it. Learning to apply broad principles in problem-solving situations is viewed as most important, with minimal guidance provided by instructors and maximum opportunity for exploration by students. In contrast, in the Guided Discovery approach, curriculum developers identify specific learning objectives for each problem, and these objectives are provided to instructors who use them to organize group discussion and student learning. These curricula can be highly structured, with careful sequencing of instructional experiences. Students may or may not be aware of the structure and the specific objectives: their experience may be quite similar to students in programs using the Open Discovery approach. In practice, the Open and Guided Discovery approaches are probably best viewed as opposite ends of a continuum. Programs vary along the continuum, and, within a program, problems (and groups) also vary.

Assessment in programs using the Open Discovery approach often focuses on process variables such as self-directedness, motivation, effort, problem-solving, and attitudes. Assessment of learning outcomes is genuinely problematic, because each student is encouraged to pursue a somewhat different course of study. Use of traditional multiple-choice tests, in particular, is often viewed as inappropriate, because they may cause students to "study to the test," thus discouraging students from self-determination of the material to be learned and the process for learning it.

Assessment of learning outcomes poses fewer problems when the Guided Discovery approach is used, since the same learning objectives that guide problem development and use can also guide test development. To achieve congruence with curricular goals, assessment should focus on students' understanding of basic mechanisms of health, disease and treatment. Well-written multiple-choice tests can play a major role in assessment, as long as they stress application of basic science

knowledge to patient care. Tests using “case clusters” — multiple-choice questions associated with the same patient presentation — are particularly appropriate for PBL courses.

An example of a simple case cluster is shown below. It consists of a brief case presentation, followed by a series of three multiple choice questions. Each question addresses a somewhat different aspect of the case, looking at the clinical situation from a variety of perspectives. Like PBL more generally, use of test material like this emphasizes learning of basic science information so that it is organized to be useful in provision of patient care.

***A 34-year-old woman has had severe watery diarrhea for the past four days. Two months earlier she had infectious mononucleosis. She abuses drugs intravenously and has antibodies to HIV in her blood. Physical examination shows dehydration and marked muscle weakness.***

1. Laboratory studies are most likely to show
  - A. decreased serum  $K^+$  concentration
  - B. decreased serum  $Ca^{2+}$  concentration
  - C. increased serum  $HCO_3^-$  concentration
  - \*D. increased serum  $Na^+$  concentration
  - E. increased serum pH
2. In evaluating the cause of the diarrhea, which of the following is most appropriate?
  - A. Colonic biopsy to identify *Giardia lamblia*
  - B. Culture of the oral cavity for *Candida albicans*
  - C. Duodenal biopsy to identify *Entamoeba histolytica*
  - D. Gastric aspirate to identify *Mycobacterium avium-intracellulare*
  - \*E. Stool specimen to identify *Cryptosporidium*
3. Further studies to evaluate her HIV infection show the ratio of helper T lymphocytes to suppressor T lymphocytes to be 0.3. This occurs because HIV
  - A. induces proliferation of helper T lymphocytes
  - B. induces proliferation of suppressor T lymphocytes
  - \*C. infects cells with CD4 receptors
  - D. infects macrophages
  - E. stimulates the synthesis of leukotriene

In addition to principles described earlier in this manual, there are two more considerations required in preparing case clusters: cueing and hinging. First, it is desirable to avoid “cueing” — providing hints at the answers to earlier questions in later questions. Students are very likely to “read ahead” for these clues, and item writers should avoid providing them. For example, in a cluster describing a patient with chest pain, if the first question addresses the most likely cause of the pain and the second requires selection of the most appropriate drug treatment, it is important that each of the diagnoses associated with the first question have a “matching” drug in the second (and vice versa); testwise examinees can rule out diagnoses (and drugs) simply by comparing the option lists.

Second, it is desirable to avoid “hinging” — creating questions where students must know the answer to one question in order to answer other questions — unless the topic to be tested is so important that the item writer is willing to have students receive either all of the points or none of the points associated with a cluster. The cluster on the next page, prepared by Drs. David Felten and Ralph Jozefowicz for the final examination in the University of Rochester first-year Neural Science course, illustrates one strategy to avoid hinging.

Each of the first three items focuses on a different aspect of the patient presentation, and students are likely to respond correctly to some and incorrectly to others, receiving “partial credit” for partial knowledge. The last item is probably slightly hinged on the preceding items, since it requires students to “put the whole picture together” in order to respond correctly, but

#### **Chapter 4. Item Content: Testing Application of Basic Science Knowledge**

*An unresponsive 58-year-old woman is brought to the emergency department after collapsing at a local shopping mall. Her family reports that she felt well that morning but developed a headache that progressively worsened while she was shopping. She has had hypertension and atrial fibrillation and is taking an antihypertensive medication and an oral anticoagulant. Her blood pressure is 220/130 mm Hg and her respiratory pattern is one of apnea alternating with hyperpnea. She responds only to noxious stimuli with extensor posturing involving the right arm and leg. Fundoscopic examination reveals papilledema involving the left optic disc. Pupils are 3.0/7.0 (R/L) with no reaction to light on the left. There is a left gaze preference. There is diffuse hyperreflexia (R > L) and Babinski's sign is present bilaterally.*

1. The dilated, unreactive left pupil is most consistent with injury to the left
  - A. optic nerve
  - B. optic tract
  - \*C. oculomotor nerve
  - D. lateral geniculate nucleus

Additional discussion of assessment in PBL courses and curricula can be found in:

Swanson DB, Case SM, and van der Vleuten CM. Strategies for student assessment. In: Boud, Feletti, eds. *The Challenge of Problem-Based Learning - Second Edition*. London: Kogan Page Ltd; 1997:269-282.

# Sample Items for the Basic Sciences

## Chapter 4. Item Content: Testing Application of Basic Science Knowledge



5. A patient with posthepatic cirrhosis develops rapid enlargement of the liver associated with deterioration of hepatic function. Serum concentration of which of the following is most likely to be abnormal?
- $\alpha_1$ -Antitrypsin
  - Carcinoembryonic antigen
  - Chorionic gonadotropin
  - \* $\alpha$ -Fetoprotein
  - Gastrin
6. The first-born infant of an Rh-negative 26-year-old woman who had two previous second trimester abortions has severe hemolysis and circulatory failure. This condition could have been prevented by treating the mother with
- anti-RhD IgG during the most recent pregnancy
  - \*anti-RhD IgG on termination of each of the first two pregnancies
  - anti-RhD IgM during the most recent pregnancy
  - anti-RhD IgM on termination of the first pregnancy
7. Laboratory tests on an edematous 35-year-old man show a normal serum concentration of complement and an increased serum concentration of cholesterol. Urinalysis shows 4+ protein, 0-5 erythrocytes/hpf, and several hyaline casts. Examination of tissue obtained on renal biopsy is most likely to show
- acute poststreptococcal (proliferative) glomerulonephritis
  - membranoproliferative glomerulonephritis
  - \*membranous glomerulonephritis
  - minimal change disease (lipoid nephrosis)
  - rapidly progressive glomerulonephritis
8. Genes on the bacterial chromosome have the following linkages in conjugal transfer: x and y, 25% of the time; y and z, 50% of the time. If the gene order is x-y-z, approximately what percentage of the time will x and z be transferred together?
- 1% of the time
  - 5% of the time
  - \*13% of the time
  - 20% of the time
  - 40% of the time

9. At a banquet, the menu included fried chicken, home-fried potatoes, peas, chocolate eclairs, and coffee. Within 2 hours, most of the diners became violently ill, with nausea, vomiting, and abdominal pain. Analysis of the contaminated food is most likely to yield large numbers of which of the following organisms?
- A. *Escherichia coli*
  - B. *Proteus mirabilis*
  - C. *Salmonella typhimurium*
  - \*D. *Staphylococcus aureus*
  - E. *Streptococcus faecalis*
10. Drug Y has a volume of distribution (Vd) of 75 L in both younger and older adult men. In younger adults, it has a clearance rate of 15 L/h, 50% of which is via the liver and 50% via the kidneys. For younger men, the maintenance regimen is 100 mg every 6 hours. Which of the following regimens will produce essentially the same steady-state concentration in an older man, whose creatinine clearance is reduced to half that of younger men, but whose hepatic function is unimpaired?
- A. 75 mg every 3 hours
  - \*B. 75 mg every 6 hours
  - C. 75 mg every 9 hours
  - D. 100 mg every 3 hours
  - E. 100 mg every 6 hours
  - F. 100 mg every 12 hours
11. A patient seen in the emergency department does not know which “heart drug” he is taking. His heart rate is greater than 80/min, and the PR and QRS intervals on an ECG are prolonged. The patient reports ringing in his ears. Which of the following drugs has the patient most likely been taking?
- A. Digoxin
  - B. Lidocaine
  - C. Phenytoin
  - D. Propranolol
  - \*E. Quinidine
12. An 8-year-old boy needs to be coaxed to go to school, and often, while there, he complains of severe headaches or stomach pain. Sometimes his mother has to take him home because of his symptoms. At night, he tries to sleep with his parents. When they insist he sleep in his own room, he says there are monsters in his closet. These findings are most consistent with which of the following diagnoses?
- A. Childhood schizophrenia
  - B. Normal concerns of latency-age children
  - \*C. Separation anxiety disorder
  - D. Socialized conduct disorder
  - E. Symbiotic psychosis



# Chapter 5

## Item Content:

### Testing Application of Clinical Science Knowledge

3

#### Methods for Assessment

Despite continued debate about the appropriateness of multiple-choice tests, all three Step exams of the USMLE continue to include multiple-choice questions (MCQs). In a quest for improved assessment instruments, the NBME has conducted continual research on other testing formats. For the past 25 years, a major focus of this research has been the Computer Based Examination (CBX) project, now termed Computer-based Case Simulations (CCS), which were introduced as a component of Step 3 in 1999. Since the mid-1970s, a second area of research has focused on standardized patients (SPs), which are scheduled for inclusion in the licensing examination sequence in 2004 or 2005. As with other forms of more “authentic assessment,” CCS and SP-based examinations appear to have significant advantages for assessment of aspects of clinical competence because they pose tasks for the examinee in a realistic way.

Other projects have focused on enhancing the multiple-choice format. As a result of test development research, MCQs today appear very different from those used in the past. For both content and psychometric reasons, true/false MCQ formats, such as K-types (multiple true/false) and C-types (A, B, Both, Neither), are no longer used on the licensure exams. While most of the questions on Step 2 have the traditional five options, both A-type questions and extended-matching questions can include as many as 26 options, pushing the examinee task to something closer to uncued free-response. Virtually every item on Step 2 provides a patient vignette that focuses on a task that is relevant to a new intern, such as determining the diagnosis or the next step in patient care. These items require interpretation and synthesis of the data that are provided; they also require application of knowledge to familiar or unfamiliar situations (depending on the experiences of the examinee). Like CCS and SP-based cases, modern multiple-choice items framed as brief patient vignettes present examinees with low-fidelity simulations of medical decision-making problems.

## General Issues Regarding What to Test

There are several tensions that influence the construction of each Step exam which may be relevant to you as you consider what to include on your exams. As dictated by the purpose statements, the USMLE examination system is designed to be an examination for general licensure; Step 2, for example, is designed to assess the application of knowledge required for beginning the first post-medical school year, regardless of specialty. The focus on content that is necessary for practice rather than explicitly on content that has been taught (if it were possible to catalog such a body of information) means that questions might be included on the exam that assess knowledge not uniformly taught in medical school. Conversely, topics taught in some medical schools might be omitted from the examination. In reality, there is a close association between what is taught and what is included on the examination, but the examination is designed to be an independent assessment of what students need to know as they begin their internship, regardless of whether or not it is taught. The analogous issue for individual schools is the extent to which it is appropriate to include content on an exam that has not been explicitly “taught.” For example, you need to decide to what extent you want to hold students responsible for independent learning beyond what has been covered in lectures.

A second tension on Step 2 arises from the focus on knowledge that is required by the general undifferentiated physician. When critics have questioned the appropriateness of a general examination in an era of specialization, we have countered by noting that the medical license is a general license, not a specialty-specific license, which we believe mandates that the examination be general. The generalist initiative has reduced comment in this area and has made us appear somewhat prescient — not something the NBME is often accused of. The analogous issue for individual schools is the extent to which clerkships should focus exclusively on “their” discipline, without an attempt to integrate across disciplines. In addition, faculty must decide on the extent to which clerkships should serve as a “pre-internship” experience for students, as opposed to providing an overview of topics that are relevant to all practicing physicians regardless of specialty.

The focus on questions that are appropriate for all graduating students, rather than those who are entering specific specialties has had a considerable influence on the content of Step 2 items. For example, the surgery committee is more likely to write items on patient management situations where immediate intervention is essential, or where it is important to triage the patient to a surgical floor rather than discharge him from the emergency department than to write questions related to surgical techniques that would not be considered essential knowledge for non-surgeons. The obstetrics and gynecology committee is less likely to write questions on the management of unusual high-risk pregnancies (even if most students have been exposed to these cases in the clerkship) than on the diagnosis of ectopic pregnancy — something every new intern, regardless of specialty, should recognize. Psychiatry items are unlikely to cover inpatient psychiatry but are more likely to assess ability to differentiate psychiatric from nonpsychiatric illnesses in patients who present with bizarre behavior. All questions

on Step 2 must be accepted by the entire discipline-specific committee; in addition, they must be accepted by an interdisci-

## **Chapter 5. Item Content: Testing Application of Clinical Science Knowledge**

Another type of question posing an inappropriate task is referred to as a “waiting room item.” In this item, the examinee is asked to select one of five patients for whom fetal karyotyping is most appropriate, almost as if the examinee is charged with performing fetal karyotyping on someone and he has only to look into the waiting room and select the patient who is most appropriate.

In contrast, the following item describes a patient and asks which study is most appropriate, a more reasonable task.

Questions assessing application of knowledge 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.

***Which of the following is an indication for fetal karyotyping in a 28-year-old woman?***

- A. Paternal age 55
- B. Fetal cystic hygroma on ultrasound exam
- C. Previous child with spina bifida
- D. Previous miscarriage of a triploid fetus
- E. Trisomy 21 in the woman's brother

***A healthy 28-year-old teacher is at 11 weeks' gestation. Family history is unremarkable except that both of her brothers have severe mental retardation, her mother died at 55 years of age of breast cancer, and her father is estranged. No family health records are available. Which of the following studies is appropriate?***

- A. Blood test for fragile X carrier status
- B. Blood test for phenylketonuria carrier status
- C. Chorionic villus sampling for Duchenne's muscular dystrophy
- D. Chorionic villus sampling for chromosome analysis
- E. Amniocentesis for "-fetoprotein

***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 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. Questions with clinical vignettes 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. The one instance in which use of a clinical vignette involves simply recall of an isolated fact is if the vignette describes a patient identical to one the student has read about before (eg, you use a patient vignette from a textbook or one discussed in class).

### ***Guidelines for Clinical Science Item Content***

- Test application of knowledge using clinical vignettes to pose medical decisions in patient care situations.
- Focus items on common or potentially catastrophic problems; avoid “zebras” and esoterica.
- Pose clinical decision-making tasks that would be expected of a successful examinee.
- Avoid clinical situations that would be handled by a (sub)specialist.

Questions should focus on specific tasks that the successful examinee must be able to undertake at the next stage of training



# Writing One-Best-Answer Items

## Constructing the Stem

The vast majority of questions should be written with a clinical vignette. The stem should begin with the presenting problem of a patient, followed by the history (including duration of signs and symptoms), physical findings, results of diagnostic studies, initial treatment, subsequent findings, etc. Vignettes may include only a subset of this information, but the information should be provided in this specified order. The stem should consist of a single, clearly formulated problem. The lead-in of the stem must pose a clear question so that the examinee can pose an answer without looking at the options. Satisfying the “cover-the-options” rule is an essential component of a good question.

**Good stem:** This stem provides sufficient information and can be answered without referring to the options.

*A 52-year-old man has had increasing dyspnea and cough productive of purulent sputum for 2 days. He has smoked one pack of cigarettes daily for 30 years. His temperature is 37.2 C (99 F). Breath sounds are distant with a few rhonchi and wheezes. His leukocyte count is 9000/mm<sup>3</sup> with a normal differential. Gram's stain of sputum shows numerous neutrophils and gram-negative diplococci. X-ray films of the chest show hyperinflation. Which of the following is the most likely diagnosis?*

**Stem testing isolated facts:** The following stem contains insufficient information; in order to answer the question, the examinee must use the options as a frame of reference.

*Which of the following is true about pseudogout?*

Patient vignettes should include some or all of the following components in the order indicated:

- Age, Gender** (eg, A 45-year-old man)
- Site of Care** (eg, comes to the emergency department)
- Presenting Complaint** (eg, because of a headache)
- Duration** (eg, that has continued for 2 days).
- Patient History (with Family History ?)**
- Physical Findings**
- +/- Results of Diagnostic Studies**
- +/- Initial Treatment, Subsequent Findings, etc.**

Make sure that your stem:

- Focuses on important concepts rather than trivial facts
- Can be answered without looking at the options
- Includes all relevant facts; no additional data should be provided in the options
- Is not “tricky” or overly complex
- Is not negatively phrased (ie, avoid using *except* or *not* in the lead-in)

## Fine Points on Item Stems

**Use of Real Patients.** We believe it is generally better not to base multiple-choice questions on “real patients,” particularly for tests aimed at students. As a general rule, real patients are too complicated, and the elements that are complicated are not necessarily those that are important for assessment. As noted earlier, we do include window dressing (ie, incidental findings), but do not include “red herrings” (ie, information that is intended to lead examinees away from the correct answer). Unfortunately, real patients often have “red herrings” among their findings.

**Use of Reference Materials.** We believe that it is appropriate to provide information in a test question if, in real life, someone would be likely to refer to a reference source to obtain the information. For example, in many instances, we believe it is appropriate to provide a table of normal laboratory values or a chart showing a recommended schedule of screening tests or immunizations. Of course, you might not just ask questions that require examinees to simply look up information in the chart provided, but you might, for example, ask about immunization of a 6-year-old child who had never been immunized.

**Use of Patient’s or Physician’s Own Words.** We generally do not believe it is useful to include the patient’s own words, particularly if the examinee task is to interpret nuances of language that might be affected by tone. On the other hand, it may be useful to ask the examinee to select the most appropriate physician response to a patient by asking the examinee to choose among options phrased as open-ended, closed, or leading questions.

**Patients Who Lie.** We believe all multiple-choice patients should tell the truth, or the physician’s interpretation of the patient’s story should be provided. Physicians use multiple cues to determine how truthful a patient is and many of these cues cannot be translated into written form. Thus, our items would describe a patient’s alcohol consumption as “The patient drinks 16 oz of beer with dinner each night” or “The patient’s description of his alcohol consumption is contradictory.” We would not write something ambiguous, such as “The patient ‘claims’ to drink only one bottle of beer each night.”

## Chapter 5. Item Content: Testing Application of Clinical Science Knowledge

## **Verbosity, Window Dressing, and Red Herrings: Do They Make a Better Test Item?**

Most educators stress the importance of writing item stems that are as short as possible, avoiding verbosity (ie, extra words), “window dressing” (ie, extraneous material), and “red herrings” (ie, information designed to mislead the examinee). Somewhat in opposition to this advice, we have emphasized use of clinical vignettes in item writing efforts. For USMLE Step 2, these vignettes consist of paragraph-length descriptions of clinical situations, generally followed by a question related to the diagnosis or next step in patient care. Such items stress application of knowledge by asking examinees to make clinical decisions, rather than to simply recall isolated facts. They are designed to reflect “real life tasks” by challenging examinees to first identify the findings that are important, then integrate those findings into a diagnosis or clinical action. Such items often require multiple steps in the thinking process. We have found that vignette items tend to have fewer technical item flaws than typical non-vignette items, presumably because vignettes follow a standard structure and pose questions that are clinically natural.

Despite these advantages, some have challenged the use of vignettes, believing that a vignette only makes an item more cumbersome by wrapping window dressing around the real question. Some advocate avoiding vignettes altogether; others advocate

The following trio of items indicate that, to the low-performing student, the items are not identical.

Each item was administered to senior medical students as part of the licensing examination. The grid under each item shows the percentage of “Hi” and “Lo” students (ie, those who performed in the top and bottom 20% of the examination overall) who selected each option. Almost all of the Hi group (99%) and the Lo group (90%) selected the correct option in the non-

## Long Vignette

*A 2-year-old black child developed swelling of his eyes and ankles over the past week. Blood pressure is 100/60 mm Hg, pulse 110/min, and respirations 28/min. In addition to swelling of his eyes and 2+ pitting edema of his ankles, he has abdominal distention with a positive fluid wave. Serum concentrations are: creatinine 0.4 mg/dL, albumin 1.4 g/dL, and cholesterol 569 mg/dL. Urinalysis shows 4+ protein and no blood. (Same lead-in and option list followed).*

	A	B	C	D	E
Hi	0	1	98	1	0
Lo	10	9	66	10	5

Although the item listed above is labeled “long vignette,” it is quite short in comparison to most items on USMLE Step 2. We believe that new graduates from medical school need to demonstrate proficiency in sorting through patient information, synthesizing the important findings, and reaching a conclusion. As a result, our items tend to have a mix of important and unimportant findings. On the other hand, we sometimes synthesize findings with a statement such as “the family history is noncontributory.”

## Writing Items Related to Physician Tasks

Each topic requires a slightly different approach to item writing. We provide some sample lead-ins and other tips to guide your item-writing efforts for each physician task.

### *Health and Health Maintenance*

Items in this topic area assess ability to evaluate risk factors, understand epidemiologic data, and apply preventive measures. Health and health maintenance items commonly fall into one of the following categories: 1) screening tests; 2) constructive interference; 3) immunizations/travel medicine; or 4) emergency intervention.

In general, begin the items with a clinical vignette that describes a patient. In addition to history and physical examination findings, these vignettes may include information about immunization history, risk factors, and family history. Information about the community may be relevant and therefore included, but the question should focus on the individual patient. Questions should NOT focus on the direct assessment of isolated facts. For example, avoid asking about the leading cause of death in some subpopulation; instead, focus on the applications of this knowledge. In asking about immunizations or screening tests, consider providing a chart of customary practices to avoid memorization of conflicting recommendations.

The following lead-ins are examples of those used in this category:

*Which of the following immunizations should be administered at this time?*

*Which of the following is the most appropriate screening test?*

*Which of the following tests would have predicted these findings?*

*Which of the following is the most appropriate intervention?*

*For which of the following conditions is the patient at greatest risk?*

*Which of the following is most likely to have prevented this condition?*

*Which of the following is the most appropriate next step in management to prevent [morbidity/mortality/disability]?*

*Which of the following should be recommended to prevent disability from this injury/condition?*

*Early treatment with which of the following is most likely to have prevented this patient's condition?*

*Supplementation with which of the following is most likely to have prevented this condition?*

*A 15-year-old boy has had two episodes of severe anaphylactic shock following bee stings. Which of the following is the most appropriate intervention?*

## Mechanisms of Disease

These items should test examinees' 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. Mechanisms questions should be framed in a clinical context.

In general, begin your mechanisms items with a clinical vignette of a patient and his/her symptoms, signs, history, lab study findings, etc. Then ask a question such as one of these:

*Which of the following is the most likely explanation for these findings?*

*Which of the following is the most likely location of the patient's lesion?*

*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?*

***A 10-year-old girl develops gross hematuria 14 days after a sore throat. She has a blood pressure of 170/100 mm Hg and 2+ pedal and pretibial edema. Serum urea nitrogen (BUN) level is 3.2 mg/dL.***

***Which of the following is the most likely cause?***

- A. Coarctation of the aorta
- B. Decreased production of endothelial-derived relaxant factor
- C. Increased production of aldosterone
- D. Increased production of catecholamine
- \*E. Intravascular volume of expansion

***A 32-year-old man has a purulent urethral discharge. A culture grows Neisseria gonorrhoeae sensitive to penicillin. One week after cessation of penicillin therapy, the patient has a recurrence of the urethral discharge. A culture again shows N. gonorrhoeae sensitive to penicillin. Both the patient and his sexual partner are HIV negative. Examination of the patient's sexual partner shows an anal fissure; urethral culture does not grow N. gonorrhoeae. Which of the following is the most likely cause of the recurrence of urethral infection?***

- A. Concurrent herpesvirus infection
- B. Emergence of bacterial resistance
- C. Inadequate treatment with penicillin
- \*D. Reinfection from the partner



## **Diagnosis**

For this category, write items that require examinees 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). As you write questions in this area, think about whether you want to assess ability to integrate knowledge across clerkships.

The classic diagnosis item begins with a patient description (including age, gender, symptoms and signs and their duration, history, physical findings on exam, findings on diagnostic and lab studies) and ends with a question:

*Which of the following is the most likely diagnosis?*

*Which of the following is the most appropriate next step in diagnosis?*

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

***A 52-year-old man has had increasing dyspnea and cough productive of purulent sputum for 2 days. He has smoked one pack of cigarettes daily for 30 years. His temperature is 37.2 C (99 F). Breath sounds are distant with a few rhonchi and wheezes. His leukocyte count is 9000/mm***

## ***Management***

These items assess principles of chronic and acute care in inpatient and outpatient settings. When writing management items,

## Writing Items on Difficult Topics

One of the common beliefs is that many topics do not lend themselves to a multiple-choice format. We have had reasonable success in generating items in many of these areas and recommend the following strategy.

1. After you identify a topic that you're having difficulty with, look through all sources of test material and select any questions on the topic that you think are acceptable.
2. Identify the key features of these items and try to develop a template that would enable faculty to write similar items.
3. For topics where no sample items are available, think about what you want to assess. Go beyond the list of topics by outlining tasks related to the topic that are essential for medical students to know (ie, each element of the list should include a verb).

To illustrate this process, the following paragraphs outline the process we used to write Step 2 questions on Medical Ethics and Jurisprudence. The content outline includes the following topics: 1) Consent and informed consent to treatment (eg, full disclosure, alternate therapies, risks and benefits); 2) Physician-patient relationship (eg, truth-telling, confidentiality, privacy, autonomy, public reporting); 3) Death and dying (eg, diagnosing death, life-support, autopsy, organ donation, euthanasia, suicide); 4) Birth-related issues (eg, prenatal diagnosis, abortion, maternal-fetal conflict); and 5) Research issues (eg, consent, placebos, conflict of interest, vulnerable populations).

Authors typically wrote questions such as: Which of the following is the definition of informed consent? Or they would describe a scenario and ask which ethical principle was illustrated by the scenario. This is a sample legal question from the past.

These are irreverently referred to as “who cares” questions.

***The legal basis for the eased restrictions on abortions in the US can be traced most closely to***

- A. *federal legislation*
- B. *a federal court ruling*
- C. *state legislations*
- D. *state court rulings*
- E. *AMA rulings*

We reviewed the item pool and decided that we were less interested in whether or not students knew the definitions; we wanted to assess whether or not examinees could apply ethical principles in their decisions related to patient care. We then convened a group of item writers who looked through model questions and generated new items for the exam. All questions involved a patient vignette and asked the student to indicate what the physician should do or asked the student to evaluate the appropriateness of the physician's actions indicated in the vignette; no questions focused on the definition of terms. The following are some sample items we wrote.

*A nurse is hospitalized for an appendectomy at the medical center where she is employed. One week after discharge, the assistant hospital administrator asks the surgeon what the final diagnosis was. Which of the following is the most appropriate response on the part of the surgeon?*

- A. *Answer, because it will expedite handling of insurance issues at the medical center*
- B. *Answer, because as an employee of the medical center the administrator has access to information about patients*
- C. *Answer, because of the possibility of spreading misinformation about the patient*
- D. *Decline to answer, because the administrator is not a medical doctor*
- \*E. *Decline to answer, because the information is confidential*

*An 8-year-old boy with acute lymphoblastic leukemia has experienced three relapses in the past 2 years. The only available treatment is experimental chemotherapy. Without treatment, the child is unlikely to survive for more than 6 weeks; with treatment, his prognosis is unknown. The parents do not want further treatment for their son and wish to take him home; the child also says he wants to go home. Which of the following is the most appropriate course of action?*

- A. *Discharge the child against medical advice*
- \*B. *Discharge the child routinely*
- C. *Petition the court for an order for treatment*
- D. *Report the parents to social services for medical neglect*



---

## **Section III**

# **Extended-Matching Items**

### 3

Section III provides information regarding Extended-Matching items. Chapter 6 presents the Extended-Matching item format where examinees are instructed to select the one-best answer. Chapter 7 presents the format where examinees are instructed to select some particular number of options, generally more than one.



---

## Chapter 6

# Extended-Matching (R-Type) Items

3

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 infection	L. Microangiopathic hemolytic anemia
F. Folate deficiency	M. Miliary tuberculosis
G. Glucose 6-phosphate dehydrogenase deficiency	N. Vitamin B <sub>12</sub> (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/mm<sup>3</sup> (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/mm<sup>3</sup>, and platelet count of 15,000/mm<sup>3</sup>.

**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 item stems 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 (or many) 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



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 rank-ordered with the other options. The stem of item #1 by itself is not clear; the item 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, not only in the clinical sciences, but also to assess knowledge in the basic sciences. 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 [eg, nerve] that is most likely to be [abnormal/defective/deficient/non-functioning].  
*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 [eg, 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

## Writing the Item Stems

Patient vignettes provide an excellent structure for stems both in the basic and the clinical sciences. In the clinical sciences, the vignette commonly provides the patient's age, gender, chief complaint, and site of care, followed by personal history, family history (if relevant), then physical examination information, then laboratory data (if provided). Depending upon the purpose of the set, vignettes can be brief, prototypic presentations or fuller descriptions that challenge examinees to identify key information. Generally, these items would include at least the patient's age, gender, chief complaint, and related history. Items to assess knowledge of the basic sciences, particularly for courses taught in the first year of medical school, might include less detailed prototypical descriptions.

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 reconstructing 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 student from the other students 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 sub-score 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 two item stems that require students to apply their basic science knowledge of microbiology to arrive at the most likely cause of each patient's illness.

- |                                       |                                      |
|---------------------------------------|--------------------------------------|
| A. Adenovirus                         | 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 psittaci</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>     |
| K. Epstein-Barr virus                 | (group A)                            |

For each patient with fever, 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  $\beta$ -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.

**Ans: T**

The following set includes two item stems. The first requires that the examinee synthesize information in order to determine a diagnosis; the second requires only recall of an isolated fact.

- |                           |              |
|---------------------------|--------------|
| A. Vitamin A              | I. Biotin    |
| B. Vitamin B <sub>1</sub> | J. Copper    |
| C. Vitamin B <sub>2</sub> | K. Folate    |
| D. Vitamin B <sub>6</sub> | L. Iodine    |
| E. Vitamin C              | M. Iron      |
| F. Vitamin D              | N. Magnesium |
| G. Vitamin E              | O. Niacin    |
| H. Vitamin K              | P. Zinc      |

*For each patient with clinical features caused by metabolic abnormalities, select the vitamin or mineral that is most likely to be involved.*

1. A 70-year-old widower has ecchymoses, perifollicular petechiae, and swelling of the gingiva. His diet consists mostly of cola and hot dogs.

**Ans: E**

2. Involved in clotting factor synthesis.

**Ans: H**

The following set includes two item stems. The first item stem requires that the examinee synthesize information to determine the diagnosis of trisomy 21; the second item provides this information. When you develop item stems, you need to decide the level of synthesis you will provide.

- |   |                              |
|---|------------------------------|
| A. Atrial septal defect                         | E. Patent ductus arteriosus  |
| B. Coarctation of the aorta                     | F. Pulmonic valve stenosis   |
| C. Complete transposition of the great arteries | G. Tetralogy of Fallot       |
| D. Endocardial cushion defect                   | H. Ventricular septal defect |

*For each patient, select the most likely congenital heart defect.*

1. A 3090-g infant is born at term. Physical examination shows the child to be hypotonic with a weak suck. There is no cyanosis. Prominent epicanthal folds, a large tongue, and small incurved 5th digits on the hands are present. There



This set poses a task that is clinically backward. The examinee is given a vaccine and asked to select the best patient for this vaccine. A more appropriate task would be for the examinee to be given a patient and asked for the most appropriate next

## Overview of the Steps for Writing Extended-Matching Items

1. **Identify the theme for the set.** The theme can be a chief complaint (eg, chest pain, fatigue), a disposition situation (eg, admission/discharge from the emergency department), a drug class (eg, antihypertensive agents, antibiotics).
2. **Write the lead-in for the set** (eg, *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 on 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 Extended-Matching Sets

## *Sample Anatomy Set*

### ***Sample Pharmacology Set***

- A. Acetaminophen
- B. Amiodarone
- C. ACE inhibitors
- D. Aspirin
- E. Atenolol
- F. Bleomycin
- G. Cytosine arabinoside
- H. Erythromycin
- I. Fentanyl
- J. Nalidixic acid
- K. Nitrofurantoin
- L. Penicillin
- M. Prednisone
- N. Procainamide
- O. Propranolol
- P. Sulfasalazine

**Sample Physiology Set — Clinical Features (Which additional findings are likely?)**

	<u>pH</u>	<u>PO<sub>2</sub> mm Hg</u>	<u>PCO<sub>2</sub> mm Hg</u>	<u>HCO<sub>3</sub> - mEq/L</u>
A.	7.15	98	33	11
B.	7.15	98	24	8
C.	7.30	56	80	38
D.	7.40	100	40	25
E.	7.50	100	33	25
F.	7.50	100	24	18
G.	7.50	56	33	25

*For each patient described below, select the most likely arterial blood gas findings.*

1. A 22-year-old man with a 3-week history of polyuria and polydipsia has had nausea, vomiting, and decreased responsiveness for the past 12 hours. Urinalysis (dipstick) shows 4+ glucose and 4+ ketones. **Ans: B**
2. A 25-year-old woman is brought to the emergency department 12 hours after a suicide attempt. She took approximately 100 500-mg aspirin tablets. **Ans: F**

### ***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 at the hips. His ery-



### **Sample Management Set: Disposition**

- A. Observe in emergency department
- B. Admit for surgery
- C. Admit for medical management
- D. Admit for endoscopy
- E. Admit for laparoscopy
- F. Order contrast studies
- G. Order MRI
- H. Order CT scan
- I. Order ultrasonography
- J. Send home with analgesics
- K. Send home for follow-up by personal physician

*For each of the following patients, select the most appropriate next step in patient care.*

*Items might describe patients with appendicitis, ectopic pregnancy, endometriosis, Crohn's disease, diverticulitis, pelvic abscess, sickle cell crisis, renal lithiasis, twisted ovarian cyst, or other problems that commonly present as emergencies. Other disposition sets might focus on "telephone triage," hospital transfer/discharge decisions, etc.*



### ***Sample Management Set: Diagnostic Testing***

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

*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 89 kg (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 grandmother was diagnosed with colon cancer at the age of 62 years. Physical examination shows no abnormalities. **Ans: E**

### ***Sample Option List for Electrolyte Abnormalities***

- A. Hypocalcemia
- B. Hypokalemia
- C. Hypomagnesemia
- D. Hyponatremia
- E. Hypercalcemia
- F. Hyperkalemia
- G. Hypermagnesemia
- H. Hybernatriemia

*For each of the following patients, select the electrolyte abnormality most likely to be present.*

***Sample Behavioral Sciences/Pediatrics Option List***



**Chapter 6. Extended-Matching (R-Type) Items**

## Steps for Organizing a Group to Write Clinical R-Sets

The following steps may be followed to utilize a group in writing clinical R-sets. Some groups have met over dinner; followed the steps below; and generated a first draft of a dozen or more items from each attendee. Others have scheduled this as a full-day off-campus retreat, with the goal of generating a pool of near-final items.

The organizer of the “item writing party” should think about how the participant’s time can best be spent. The yield will be substantially greater in terms of both quality and quantity if some work is done in advance. For example, there will be significantly smaller yield if the participants are expected to decide what topics to write on; time will be saved if the topics for each set are defined in advance of the meeting (Step #1 below). Similarly, more items will be developed if a draft of the options for each set is developed in advance (Step #3 below). More items will also be developed if a sample item is written as a model for each set (Step #4 below). There are situations where it is best to allow item writers considerable flexibility in determining what to write. In these circumstances, providing one or more option sets and allowing them to generate one or more options sets on their own might be workable.

Decisions will also need to be made about the composition of the pairs of item writers. In some circumstances, it might be best to allow participants to self-select their partners. In other circumstances, it might be best to assign the pairs. We have had the most success assembling item-writing pairs with similar interests, but including individuals with more diverse expertise as reviewers. If the goal is to generate interdisciplinary items for an exam, the same list of options may be provided to

- 1. Define the content domain of the exam.** For example, in developing an exam to test the ability to diagnose common clinical problems, you might define the domain by a list of the chief complaints. Time will be saved if this is done in advance of the item-writing meeting.

Abdominal mass  
Abdominal pain  
Anemia  
Ascites

Dizziness  
Easy bruising  
Fatigue  
Fever

Lymphadenopathy  
Movement abnormalities  
Nausea  
Palpitations

**4. Stress the following guidelines for writing stems.**

Each item should describe a patient with one of the diagnoses in the option list, beginning with the patient's age, gender, chief complaint, and site of care, followed by personal history, family history (if relevant), then physical examination information, then laboratory data (if provided).

Depending upon the purpose of the set, vignettes can be brief prototypic presentations or fuller descriptions that challenge examinees to identify key information.

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, include it in all items; if laboratory data are included for one item, include them in all items.

**5. Merge the pairs into a larger group to review the items.** One approach is to have the author read the item aloud; others attempt to provide the correct answer. The group reviews the option list and modifies the item or the option list to eliminate any ambiguity. Other approaches are outlined above.

**6. Type, edit, and subject the items to external review.** Items should be reviewed without the correct answer indicated after they are in their final form.

**7. Construct the test.** Select a sample of items from each complaint; save the remaining items for subsequent exams. Items can be converted into one-best-answer items by adding a lead-in and the best five (or more) options from the option list.

# Form for Writing R-Sets

Theme: \_\_\_\_\_  
(eg, a presenting complaint)

Lead-In: \_\_\_\_\_  
(eg, For each patient with fever, select the most likely diagnosis.)

## Options

A	N
B	O
C	P
D	Q
E	R
F	S
G	T
H	U
I	V
J	W
K	X
L	Y
M	Z

(Write items on separate pages)

## Sample SPSSX Code to Score Multiple-Choice Tests, including Extended-Matching Items

The following SPSSX code can be used as a model for scoring a hypothetical test including up to 100 multiple-choice items; each item may have up to 26 options but only one correct answer. It is straightforward to alter the code for any test length.

It is assumed that:

- the answer key is in a file named KEY.DAT (format: an eight-character exam ID, followed by a space and the 100 correct answers);
- the examinee responses are in a file named RESPONSE.DAT (format: social security number or any nine-digit examinee ID, a space, the eight-character exam ID, a space, the examinee's responses to the 100 items);
- the item analysis output shows the distribution of responses to each item, plus a reliability coefficient (coefficient alpha); and
- score reports with the examinee ID number, plus percent correct and standard scores (placed in a file named REPORT.LIS). REPORT.LIS can be imported into a word processor to enhance its appearance.

```

TITLE                SAMPLE SPSSX SETUP FOR SCORING A HYPOTHETICAL 100-ITEM
SUBTITLE             MULTIPLE CHOICE TEST — DAVE SWANSON, SEPTEMBER 6, 1991
SET                  LENGTH=64/WIDTH=132

COMMENT              READ IN AND SAVE THE ANSWER KEY
FILE HANDLE          KEYDAT/NAME='KEY.DAT'
DATA LIST            FILE=KEYDAT/
                    EXAMCODE,KEY1 TO KEY100
                    (A8,1X,100A1)

FILE HANDLE          KEYSYS/NAME='KEY.SYS'
SAVE                OUTFILE=KEYSYS

COMMENT              READ IN EXAMINEE RESPONSE STRINGS
FILE HANDLE          RESPDAT/NAME='RESPONSE.DAT'
DATA LIST            FILE=RESP/
                    SSN,EXAMCODE,RESP1 TO RESP100
                    (F9.0,1X,A8,1X,100A1)

COMMENT              ADD THE ANSWER KEY TO RESPONSE STRING RECORDS
MATCH FILES          FILE=*/TABLE=KEYSYS/BY EXAMCODE

COMMENT              COMPARE THE KEY TO RESPONSES AND CREATE A 0/1 VECTOR OF
COMMENT              INCORRECT/CORRECT ANSWERS
VECTOR              SCORE(100,F1.0)
DO REPEAT            K=KEY1 TO KEY100/R=RESP1 TO RESP100/S=SCORE1 TO SCORE100
COMPUTE              S=0
IF                   (K EQ R) S=1
END REPEAT

COMMENT              PRINT A CROSSTABULATION OF RESPONSES FOR EACH ITEM
TABLES              FORMAT=CWIDTH(10,3) NSPACE LIGHT/
                   TABLE= RESP1 +
                   RESP2 +
                   ... + (you'd actually need to type all of these in)
                   RESP100 BY (LABELS)/

                   VECTOR              SCOOVECTOR( 5T61IB PRNLF6 0 1ED3 9 U*FT WLARI1 TDESP100/S=SCORE1 TO

COMMENT              COMPUTESP100/S=SCORE1Q00,F1.ODE

COMMENT              FILEHANDLERESPDAMEANSD.TMP'KEY.SYS'CO0SAV-13349.5(MEANSD(BE#YSYS/BYLE=RESP/) T 15. 5PCMEAN=MEAN(PCORE1/LPCOD=SD(PCORE1(,

```

## Chapter 6. Extended-Matching (R-Type) Items



## Comparison of Items in Five-Option and Extended-Matching Format

In several studies to investigate the optimal number of options for multiple-choice items, we have consistently found that, other things being equal, more options are better than fewer options. Based on items used on NBME exams, extended-matching items are more discriminating than any other format; 5-option A-types are second best; and the various forms of true/false items are the worst. In controlled studies comparing otherwise equivalent 5-option and extended-matching items, extended-matching items were found to be more discriminating than 5-option items; comparable levels of reproducibility can be achieved with the extended-matching format using one-third fewer items than with 5-option items. Extended-matching items were also found to be more difficult than content-parallel 5-option items: there is a lower probability of guessing the correct answer, and item writers are not always able to select the most functional distractors in reducing the number of options to five.

The following table shows examinee responses to a sample item presented in a 5-option format and in a 15-option format. The item was originally written as a 15-option item; the item writer then reduced the number of options to five by selecting what he thought were the best distractors (B, F, G, J, N). The item was markedly easier in the 5-option format (p value of 81 vs 59), and the discrimination was markedly lower (not shown). There is an increased probability of examinees selecting the correct answer in the 5-option format, especially because item writers do not uniformly identify the most salient distractors (eg, option D). Tests constructed of extended-matching items tend to spread out the lower ability students; the extended list of options gives them more opportunity to show what they don't know.

Format	Examinee Responses														
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5-option		81		*		4	0			14			*	1	
15-option	2	59	2	13	1	2	1	0	1	7	2	2	8	0	0

For additional information, see also:

Case SM, Swanson DB, Ripkey DR. Comparison of items in five-option and extended-matching format for assessment of diagnostic skills. *Academic Medicine*. 1994;69(suppl):S1-S3.

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

Swanson DB, Case SM. Trends in written assessment: a strangely biased perspective. In: Harden R, Hart I, Mulholland H, eds. *Approaches to Assessment of Clinical Competence*. Norwich, England: Page Brothers; 1992:38-53.

---





---

## Chapter 7

# Pick N Items: An Extension of the Extended-Matching Format

3

The Pick N format may be similar to either the extended-matching or the A-type format; the primary difference is that the examinee is told to pick 2, 3, 4, or even 5 of the options listed. As with extended-matching sets, the option list may include up to 26 options. The format was developed to replace negative items or items with double options, particularly in areas such as Health Maintenance and Disease Prevention. Items might focus on various patients with different risk factors who come for routine examination; the examinee would be asked to select the laboratory studies or immunizations that should be ordered for each patient. The format could also be used for emergency management items where several procedures would be carried out simultaneously; the examinee would be asked to select a specific number of actions from the option list.

The item-writing rules are the same as for extended-matching sets. The options should be short (usually a single word or very short phrase); the patient vignettes can be long. If the set asks for management decisions, each vignette should contain all relevant history and physical examination data. As with extended-matching items, the format works well for items that appear to be extremely easy; tricky or unnecessarily complex vignettes should be avoided.

The Pick N format is designed to specify exactly how many options to select. The rationale for this decision is derived from the essential difference between true/false and one-best-answer items, where true/false items require the examinee to indicate all responses that are appropriate, and one-best-answer items require the examinee to indicate a specific number of responses. Specifying the number of options to be selected changes the task from a true/false task to a best-answer task.

Research indicates that partial credit scoring is preferred but logistical considerations might preclude this. If you use all-or-nothing scoring, the items may be extremely difficult, and it is best to require examinees to select only two or three options, rather than more.

For additional information, see also:

Ripkey DR, Case SM, Swanson DB. A “new” item format for assessing aspects of clinical competence. *Academic Medicine*. 1996;71 (suppl):S34-S36.

- |                               |                              |
|-------------------------------|------------------------------|
| A. Diabetic polyneuropathy    | F. Multiple sclerosis        |
| B. Huntington's disease       | G. Parkinson's disease       |
| C. Lateral medullary syndrome | H. Pontine glioma            |
| D. Lead encephalopathy        | I. Tabes dorsalis            |
| E. Medulloblastoma            | J. Wernicke's encephalopathy |

*A 50-year-old man has the gradual onset of mental confusion, disorientation, and loss of short-term memory. He has a left footdrop. A blood smear shows microcytosis and basophilic stippling of erythrocytes. (Select the two most likely diagnoses)*

In the example above, there would be disagreement about what diagnoses are likely, but the task becomes clear if the examinee is told to select the two most likely diagnoses. The options can be diagnosed as follows.

<b>F</b>	<b>C</b>	<b>E</b>	<b>D</b>	<b>A</b>	<b>G</b>	<b>H</b>	<b>B</b>
<i>Least Likely Diagnosis</i>						<i>Most Likely Diagnosis</i>	

### Sample Pick N Set

- |                                      |   |
|--------------------------------------|---|
| A. Calcium                           | G. Vitamin B <sub>6</sub>                   |
| B. Fluoride                          | H. Vitamin B <sub>12</sub> (cyanocobalamin) |
| C. Folic acid                        | I. Vitamin C                                |
| D. Iron                              | J. Vitamin D                                |
| E. Vitamin A                         | K. Vitamin E                                |
| F. Vitamin B <sub>1</sub> (thiamine) |   |

*For each child, select the appropriate vitamin or mineral supplements.*

1. A 1-month-old infant is brought to the physician for a well-child examination. He has been exclusively breast-fed, and examination shows normal findings. (SELECT 2 SUPPLEMENTS). **Ans: B, J**
2. A 6-year-old girl has cystic fibrosis. She has been taking no medications. (SELECT 3 SUPPLEMENTS). **Ans: E, J, K**

### Sample Pick N Set

- |  |                              |
|--|------------------------------|
| A. Analysis and culture of cerebrospinal fluid | F. Urinalysis                |
| B. Blood culture                               | G. Urine culture             |
| C. Complete blood count                        | H. X-ray film of the abdomen |
| D. Examination of the stool for leukocytes     | I. X-ray film of the chest   |
| E. Measurement of serum electrolyte levels     |                              |

*For each child with fever, select the appropriate initial diagnostic studies.*

1. A previously healthy 1-year-old girl is brought to the emergency department because of fever for 1 day. Her temperature is 41 C (105.8 F). She is otherwise asymptomatic. Physical examination shows no abnormalities. (SELECT 4 STUDIES). **Ans: B, C, G, I**
2. A previously healthy 10-day-old newborn is brought to the emergency department because of fever for 2 hours. He was born at term after an uncomplicated pregnancy. His temperature is 39 C (102.2 F). Physical examination shows no abnormalities. (SELECT 6 STUDIES). **Ans: A, B, C, E, G, I**
3. A 7-year-old boy with sickle cell disease is brought to the emergency department because of fever for 1 day and chest pain for 1 hour. His temperature is 39.5 C (103.1 F). Breath sounds are slightly decreased in the right lower lung; he is not in respiratory distress. (SELECT 3 STUDIES). **Ans: B, C, I**

**PATIENT CHART****Patient History**

**Sex:** male  
**Current age:** 28 years  
**Chief complaint:** health maintenance examination  
**Social history:**  
**Marital status:** single  
**Occupation:** computer programmer  
**Alcohol:** 2-4 beers/weekend  
**Smoking:** 5-10 cigarettes daily from age 16-24  
**Exercise prgm:** sedentary  
**Medical history:**  
**Childhood:** obese since grade school  
**Immunizations:** all childhood immunizations; last tetanus toxoid age 15;  
no immunizations since childhood  
**Screening:** no physician visits since college  
**Family history:**  
**Parents:** father age 57; hypertensive mother age 55; obese and hypertensive  
**Siblings:** none  
**Children:** none  
**Current medications:** none  
**Allergies:** none

**Physical Examination**

**Height:** 178 cm (70 in)  
**Weight:** 134 kg (295 lb)  
**Vital signs:**  
**Blood pressure** 148/86 mm Hg  
**Pulse** 90/min  
**Respirations** 16/min  
**Skin:** erythematous rash in groin  
**Abdominal:** obese  
**Laboratory studies:** none ordered

1. For the patient whose chart is shown, select the conditions for which he is at increased risk.  
(SELECT 4 CONDITIONS)





---

## **Section IV**

### **Additional Issues**

3

This section includes some additional issues that are related to testing.



---

## Chapter 8

# Interpretation of Item Analysis Results

3

Many schools provide faculty with item analysis output following each multiple-choice examination. This output is an excellent source of information about an item and is useful in evaluating the quality of the item, as well as in evaluating the accuracy of the answer key.

The following are sample results from four items; each illustrates a common situation. The students taking the test were divided into a Hi group and a Lo group, based on their performance on the total test. If you have a small number of examinees, include the top 50% of the students in the Hi group and the bottom 50% in the Lo group. If you have a large number of examinees, you might include the top 25% in the Hi group and the bottom 25% in the Lo group.

Typically, item analysis output indicates the percentage of students in each group who selected each option. Often, it also includes some measure of item difficulty (eg, the “p-value” or the proportion of students who answered the item correctly) and some measure of discrimination (eg, a biserial or a point biserial). We recommend that attention be focused on the pattern of responses rather than on the difficulty level or discrimination index.



**Item #3**

**Chapter 8. Interpretation of Item Analysis Results**



---

## Chapter 9

# Establishing a Pass/Fail Standard

3

### Definitions and Basic Principles

Standards may be classified as either *relative* or *absolute*. A *relative standard* is based on the performance of the group taking the test. Examinees pass or fail depending upon how well they perform relative to other examinees taking the test. The following are examples of *relative* standards:

Those scoring below 1.2 standard deviations below the mean will fail.  
The bottom 20 percent of the group will fail.

In contrast, an *absolute standard* does not compare the performance of one examinee with the others who are taking the test. Examinees pass or fail based only upon how well they perform, regardless of the performance of other examinees. All examinees could pass or all could fail. The following is an example of an *absolute* standard:

Those answering less than 60 percent of the questions correctly will fail.

Unless there are strong reasons to fail a given number of examinees, an absolute standard (based on examinee performance) is preferred over a relative standard (based on a particular failure rate).

#### *Basic Principles of Setting Standards*

- Regardless of the procedure used, setting standards requires judgement. Setting standards will always be arbitrary, but need not be capricious.
- Unless there is a specific reason to fail a given number of examinees (eg, there are only a fixed number of slots available), a standard based on examinee mastery of exam content is preferred over a standard based on a particular failure rate.



- It is wise to involve multiple informed judges in the standard-setting process. Differences of opinion will occur, and use of multiple judges will reduce hawk/dove effects.
- Judges should be provided with data on examinee performance at some point in setting standards. Setting standards without such data may lead to uninformed standards and unreasonable results.

A helpful “how-to” reference on standard setting is:

Livingston SA, Zieky MJ. *Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests*. Princeton, NJ: Educational Testing Service; 1982.

## Two Standard-Setting Methods Based on Judgements about Items

### *The Modified Ebel Procedure*

- A group discusses the characteristics of the “borderline examinee”: an examinee whose skills are just good enough to allow him/her to pass.
- Judges categorize items as “Essential,” “Important,” or “Indicated.”
- Judges indicate the number of items in each category that a borderline examinee would obtain.
- The pass/fail standard is calculated as the percentage of possible points that a borderline examinee would obtain.

***The Modified Angoff Procedure***



# Revisiting Compromise Standards: The Hofstee Method

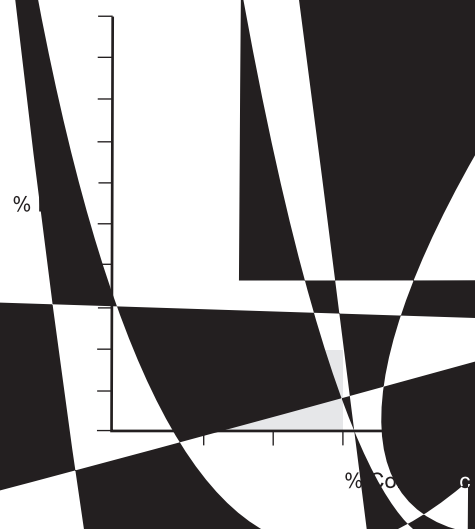
More "compromise models" have been developed that utilize the same basic principles. One of these methods, the Hofstee method, is described below.

1. Review a copy of the exam.
2. Determine the following values, which define acceptable standards:
  - a. Minimum acceptable percentage of failing examinees (minimum failure rate)
  - b. Maximum acceptable percentage of failing examinees (maximum failure rate)
  - c. Minimum acceptable percentage which would allow someone to pass (minimum passing point)
  - d. Maximum acceptable percentage required for someone to pass (maximum passing point)

3. Assume that a normal distribution curve showing the failure rate versus the passing score is plotted. (In the figure, the horizontal axis extends from bottom left to top right.)

4. The minimum and maximum values in #2 are drawn, forming a rectangle. The minimum and maximum values of the group of examinees are also drawn. For example, the appropriate minimum failure rate would be between 0 and 20% (see horizontal lines); the appropriate maximum failure rate would be between 50 and 60% (see vertical lines).

5. A line is drawn on the diagonal from the top left to the bottom right. The point where this line intersects the curve is the standard (ie, just above the minimum passing point in the figure).



A useful reference on the Hofstee method for determining examination standards. *Journal of Educational Measurement*;22:263-269.

---

## Chapter 10

# Miscellaneous Thoughts on Topics Related to Testing

3

*Comments on a hodge-podge of topics related to testing are provided below. In general, the points made are speculative and based on anecdotal experience rather than evidence. That is, they reflect our biases rather than the results of research.*

### **Multiple Station Exams (a.k.a. Practical Exams, Steeplechases, OSCEs)**

Though logistically complex to set up and administer, these are very useful in the basic sciences, particularly to assess hands-on skills that cannot be measured with paper-and-pencil tests (eg, ability to use a microscope, to perform a laboratory procedure). In addition, reproduction of some kinds of material (eg, results of imaging studies, color pictorial material) is very expensive; in such situations, the multiple-stations approach can be used to reduce test administration costs.

### **Take-Home Exams**

Take-home exams can be a substantial learning experience for students by stimulating them to read broadly and deeply on important topics. Unfortunately, students tend to produce tomes as answers, and it can be unclear if submitted answers represent the student's own work. The same advantages can be gained by distributing (a superset of) test questions in advance, and administering (a subset of) questions as a timed test.

### **Open-Book Tests**

Open-book tests can be a very good idea because of the impact on the kinds of questions that faculty prepare. For open-book tests, it is pointless to ask questions about isolated facts that can be looked up quickly on a single page of a text book, so test material developed for these tests tends to focus more on understanding of key concepts and principles in problem situations.

### **Frequent Short Quizzes versus Infrequent Tests**

Infrequent testing makes each exam a major event; students may even stop attending class to prepare, and this seems undesirable. In addition, with infrequent tests, students may be unable to determine if they are studying the right material or learn-

---

# **Appendix A**

## **The Graveyard of NBME Item Formats**







---

## D-Type Items

D-type items were complex matching items in which each item consisted of three functional disturbances (designated by a letter) and five situations (in a numbered list). The examinee was instructed to 1) select the functional disturbance or category that four of the five situations were related to and 2) indicate the one situation that did not belong in that category. It was believed that these items required discriminatory understanding of a number of similar factors. However, D-type items were difficult to write, and the directions were confusing. In addition, they did not discriminate between knowledgeable and unknowledgeable examinees.

### *Sample D-type Item*

DIRECTIONS: There are two responses to be made to *each* of the following questions. In the left-hand list are three lettered categories. Exactly four of the five numbered items in the right-hand list are related in some way to ONE of these categories. (1) on the appropriate line in the answer sheet blacken the space under the letter of the category in which these four items belong. (2) Then blacken the space under the number of the item in the right-hand list that does NOT belong in the same category with the other four.

- A. Eosinophilia of diagnostic significance
- B. Plasmacytosis of diagnostic significance

- 1. Trichinosis
- 2. Multiple myeloma

## K-Type Items

K-type items were the most commonly used multiple true/false item format at the National Board. They consisted of a stem followed by four options, one or more of which was correct. It was believed that K-type items tested in-depth knowledge or understanding of several aspects of a disease, a process, or a procedure, and required an examinee to be familiar with several different facts about a given topic. However, K-type items were criticized as being too complicated, requiring the examinee to constantly keep the answer code in mind. In addition, the possible response combinations introduced a cueing effect that reduced item discrimination and lowered test reliability. It was difficult to write good, unambiguous true/false items. Because the items could include only absolutely true or false facts, K-type items could not be used to assess clinical judgment except in comparisons (eg, "Drug X is better than Drug Y in treating disease K"). K-type items were more difficult and less discriminating than other item types. In addition, they were less efficient than other MCQ formats, and the relative reliability per unit of test time was lower.

### Sample K-type Item

<b><u>Directions Summarized</u></b>				
<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>
1, 2, 3 only	1, 3 only	2, 4 only	4 only	All are correct

1. A child suffering from an acute exacerbation of rheumatic fever usually has
  - (1) an elevated sedimentation rate
  - (2) a prolonged PR interval
  - (3) an elevated antistreptolysin O titer
  - (4) subcutaneous nodules



## E-Type Items

E-type items were multiple true/false items that are based on the analysis of relationships. Examinees who took E-type items still refer to them as the “True, True and Unrelated” items. The E-type consisted of a sentence with two main parts: an assertion and a reason for that assertion. The examinee was directed to select A if both were true statements and the reason was a correct explanation of the assertion; B if both were true statements but the reason was not a correct explanation of the assertion; C if the assertion was true but the reason was a false statement; D if the assertion was false but the reason was a true statement; E if both assertion and reason were false statements. It was thought that good reasoning skills and an understanding of the basic principles were required to answer this item type correctly. However, E-type items were difficult to construct, and examinees found them to be confusing.

### *Sample E-type Item*

<b>Directions Summarized</b>			
<b>A</b>	True	True	Reason is a correct explanation.
<b>B</b>	True	True	Reason is NOT a correct explanation.
<b>C</b>	True	False	
<b>D</b>	False	True	
<b>E</b>	False	False	

#### **Assertion**

#### **Reason**

- |  |         |   |
|--|---------|---|
| 1. Herpes simplex is usually regarded as an autogenous infection | BECAUSE | patients given fever therapy frequently develop herpes. |
| 2. Cow's milk is preferable to breast milk in infant feeding     | BECAUSE | cow's milk has a higher content of calcium.             |



## I-Type Items

The I-type item was similar to the H-type. It consisted of pairs of phrases that describe conditions or quantities that might vary in relation to each other. The examinee was directed to select A if the two phrases were related directly (ie, an increase in the first was accompanied by an increase in the second or a decrease in the first was accompanied by a decrease in the second); B if the phrases were related inversely (ie, an increase in the first was accompanied by a decrease in the second or a decrease in the first was accompanied by an increase in the second); or C if the changes were independent of one another.

### *Sample I-type Item*

**DIRECTIONS:** Each of the following pairs of phrases describe conditions or quantities that may or may not be related. On the appropriate line of the answer sheet blacken the space under

- A** if increase in the first is accompanied by increase in the second or if decrease in the first is accompanied by decrease in the second
- B** if increase in the first is accompanied by decrease in the second or if decrease in the first is accompanied by increase in the second
- C** if changes in the first are not necessarily accompanied by changes in the second.

1. (A) Urine volume  
(B) Urine specific gravity
2. (A) Plasma protein concentration  
(B) Colloid osmotic pressure of plasma

Neither the H- nor I-type formats were particularly popular. Because there were fewer options than in other item types, there was an increased chance of guessing the correct answer. In addition, the items tended to focus on minor details rather than scientific concepts.

In his series in the Federation Bulletin, Morton (1985-86) implied that different item types were included on medical licensure examinations simply to add variety to a lengthy examination. But, 25 years after the National Board converted from an





---

**Appendix B**  
**Sample Item-Writing Templates, Items,**  
**Lead-Ins, and Option Lists**  
**For the Basic and Clinical Sciences**



---

## Appendix B

### Sample Item-Writing Templates, Items, Lead-Ins, and Option Lists for the Basic and Clinical Sciences

3

#### Gross Anatomy

##### Template for Gross Anatomy Items

Item Stem (Patient Vignette): (Describe a patient with a problem)

Lead-In: A defect is most likely to be present in which of the following structures/processes?

Options: (List of structures/processes)

A 65-year-old man has difficulty rising from a seated position and straightening his trunk, but he has no difficulty flexing his leg. Which of the following is most likely to be injured?

- A.\* Gluteus maximus
- B. Gluteus minimus
- C. Hamstrings
- D. Iliopsoas
- E. Obturator internus

A 30-year-old man has loss of pain and temperature sensation on the left side of the face and on the right side of the body from the neck down; partial paralysis of the soft palate, larynx, and pharynx on the left; and ataxia on the left. This syndrome is most likely to result from thrombosis of which of the following arteries?

- A. Basilar
- B. \*Right posterior inferior cerebellar
- C. Left posterior inferior cerebellar
- D. Right superior cerebellar
- E. Left superior cerebellar

*Insert patient vignette describing a patient with a problem.* Which of the following is the most likely disorder of the knee?

- A. Chondromalacia patellae
- B. Dislocation (tibiofemoral)
- C. Fracture of the patella
- D. Ganglion cyst
- E. Giant cell tumor of bone
- F. Osteochondritis dissecans
- G. Osteosarcoma
- H. Prepatellar bursitis
- I. Septic arthritis
- J. Tear of the meniscus

*Insert patient vignette describing a patient with a problem.* Which of the following is the vessel into which contrast medium should be injected during fluoroscopy to visualize the site of the abnormality?

- A. Celiac artery
- B. Internal iliac artery
- C. Inferior mesenteric artery
- D. Superior mesenteric artery
- E. Renal artery
- F. Portal vein

*Insert patient vignette describing a patient with a problem.* Which of the following gastrointestinal disorders is the most likely cause of these findings?

- A. Candidal esophagitis
- B. Diverticulitis
- C. Hiatal hernia
- D. Peptic ulcer
- E. Pseudomembranous enterocolitis
- F. Pyloric stenosis
- G. Regional enteritis
- H. Subphrenic abscess
- I. Ulcerative colitis

**Isolated Fact Version:**

Which of the following areas is supplied with blood by the posterior inferior cerebellar artery?

**Application of Knowledge Version:**

A 62-year-old man develops left-sided limb ataxia, Horner's syndrome, nystagmus, and loss of pain and temperature sensations on the face. Which of the following arteries is most likely to be occluded?

## Behavioral Sciences

*Insert patient vignette describing a child's age and what he can do.* Which of the following best describes the level of development?

<u>Cognitive</u>	<u>Gross Motor</u>	<u>Social</u>
<u>Language Skills</u>	<u>Skill</u>	<u>Skills</u>
A. Normal	Normal	Normal
B. Normal	Normal	Delayed
C. Normal	Delayed	Normal
D. Normal	Delayed	Delayed
E. Delayed	Normal	Normal

A 35-year-old woman comes to the physician after finding a lump in her breast. She is crying and tells the physician that her mother died of breast cancer. Which of the following responses by the physician is the most appropriate?

- A. "I can see finding the lump was upsetting to you."
- B. "I doubt that this lump is breast cancer."
- C. "Tell me about how finding the lump made you feel."
- D. "Tell me more about the lump."
- E. "There is no reason to worry until we biopsy the lump."
- F. "Treatment for breast cancer has improved a lot since your mother died of breast cancer."

**Isolated Fact Version:**

**Application of Knowledge Version:**

## Biochemistry

*Insert patient vignette describing a patient with a problem.* Which of the following digestive enzymes or cofactors is most likely to be involved?

- A. Amylase
- B. Chymotrypsin
- C. Colipase
- D. Enterokinase
- E. Lactase
- F. Lipase
- G. Pepsin
- H. Sucrase
- I. Trypsin

*Insert patient vignette describing a patient with a problem.* Which of the following laboratory findings is the most likely?

- A. Hypercalcemia
- B. Hypocalcemia
- E. Hypernatremia
- F. Hyponatremia



**Non-Vignette Version:**

An inherited metabolic disorder of carbohydrate metabolism is characterized by an abnormally increased concentration of hepatic glycogen with normal structure and no detectable increase in serum glucose concentration after oral administration of fructose. These two observations suggest that the disease is the result of the absence of which of the following enzymes?

- A. Fructokinase

**Vignette Version:**

A 6-month-old infant has an enlarged liver. Evaluation for metabolic disease shows an abnormally increased concentration of hepatic glycogen with normal structure and no detectable increase in serum glucose concentration after oral administration of fructose. These two observations suggest that the disease is a result of the absence of which of the following enzymes?

- A. Fructokinase
- B. Glucokinase



## Genetics

A male neonate has severe hemolysis and circulatory failure. The 26-year-old mother is Rh negative and had two previous second-trimester abortions. This condition could have been prevented by administration of which of the following to the mother?

- A. Anti-RhD IgG during the most recent pregnancy
- B.\* Anti-RhD IgG on termination of each of the first two pregnancies
- C. Anti-RhD IgM during the most recent pregnancy
- D. Anti-RhD IgM on termination of the first pregnancy

Genes on the bacterial chromosome have the following linkages in conjugal transfer: x and y, 25% of the time; y and z, 50% of the time. If the gene order is x-y-z, approximately what percentage of the time will x and z be transferred together?

- A. 1%
- B. 5%
- C. 13%
- D. 20%
- E.\* 40%

Hereditary hyperammonemia, characterized by a grossly abnormal EEG and an increased blood ammonium concentration, is most likely to be caused by a deficiency of which of the following enzymes?

- A. Asparagine synthetase
- B.\* Carbamoyl phosphate synthetase I
- C. Fumarase
- D. Glutamate-oxaloacetate aminotransferase
- E. Glutaminase

## Histology/Cell Biology

*Insert description of "action".* Which of the following cellular organelles is most directly involved?

- A. Golgi complex
- B. Lysosome
- C. Peroxisome
- D. Mitochondria
- E. Nuclear envelope
- F. Secretory (zymogenic) granule
- G. Rough endoplasmic reticulum
- H. Smooth endoplasmic reticulum

## Lab Vignette

Several contiguous cells are labeled with a fluorescent dye that cannot cross cell membranes. One cell is experimentally bleached with a light that destroys the dye but soon recovers dye fluorescence. This recovery is best explained by the presence of which of the following structures between the bleached cell and its fluorescent neighbors?

- A. Basal lamina
- B. Desmosomes (maculae adherentes)
- C.\* Gap junctions
- D. Glycosaminoglycans
- E. Tight junctions (zonulae occludentes)

# Microbiology

## Template for Microbiology Items:

Item Stem (Patient Vignette): (Describe a patient with a problem)

Lead-In: Infection with which of the following organisms is most likely?

Options: (List of pathogens)

At a banquet, the menu included fried chicken, home fried potatoes, peas, chocolate eclairs, and coffee. Within 2 hours, most of the diners became violently ill, with nausea, vomiting, and abdominal pain. Analysis of the contaminated food is most likely to yield large numbers of which of the following organisms?

- A. *Escherichia coli*
- B. *Proteus mirabilis*
- C. *Salmonella typhimurium*
- D.\* *Staphylococcus aureus*
- E. *Streptococcus faecalis*
- F. *Enterococcus*

Insert patient vignette describing a person with a problem. Which of the following toxins is most likely to be involved in pathogenesis?

- A. Botulinum toxin
- B. Diphtheria toxin
- C. Pertussis toxin
- D. Shiga toxin
- E. Tetanus toxin
- F. Cholera enterotoxin
- G. *Clostridium difficile* enterotoxin
- H. *Escherichia coli* heat-stable enterotoxin
- I. *Staphylococcus aureus* enterotoxin

*Insert patient vignette describing a patient with a problem.* Which of the following is the most appropriate therapeutic agent?

- A. Acyclovir
- B. Amphotericin B
- C. Erythromycin
- H. Pyrantel pamoate
- I. Pyrazinamide
- J. Rifampin

### Sample Item Set

A 2-year-old boy has had recurrent infections since he was 6 months old. Serum complement concentrations, phagocytic function, and bactericidal of neutrophils are normal. A skin test with *Candida* antigens results in 2 cm of induration by 48 hours. Which of the following provides the best explanation for the time of onset of the disease at 6 months of age?

- A. Development of a viral infection
- B. Exposure to rare bacteria
- C. Loss of passive immunity from the mother
- D. A maturation defect in the thymus
- E. Transplacental transfer of IgM antibodies

Which of the following laboratory tests is most likely to be abnormal?

- A. Activity of myeloperoxidase
- B. CD4/CD8 T lymphocyte ratio
- C. Sweat chloride concentration
- D. Fc receptors on macrophages
- E. Serum IgG concentration

## Neuroscience (Neuroanatomy and Neuropathology)

### Template for Neuroanatomy Items

Item Stem (Patient Vignette): (Describe a patient with a problem)

Lead-In: An abnormality is most likely to be present at which of the following locations?

Options: (List of sites)

*Insert patient vignette describing a patient with a problem.* Which of the following cranial nerves is the most likely site of the underlying lesion?

- |                     |                            |
|---------------------|----------------------------|
| A. Olfactory nerve  | G. Facial nerve            |
| B. Optic nerve      | H. Vestibulocochlear nerve |
| C. Oculomotor nerve | I. Glossopharyngeal nerve  |
| D. Trochlear nerve  | J. Vagus nerve             |
| E. Trigeminal nerve | K. Spinal accessory nerve  |
| F. Abducens nerve   | L. Hypoglossal nerve       |

*Insert patient vignette describing a patient with a neurologic abnormality.* Which of the following branches of the brachial plexus is most likely to be affected?

- |                      |                     |
|----------------------|---------------------|
| A. Axillary          | G. Musculocutaneous |
| B. Dorsal scapular   | H. Radial           |
| C. Suprascapular     | I. Long thoracic    |
| D. Upper subscapular | J. Thoracodorsal    |
| E. Lower subscapular | K. Ulnar            |
| F. Median            |                     |



*Insert patient vignette describing a patient with a problem.* Which of the following components of the motor system is most likely to be involved?

- A. Basal ganglia
- B. Cerebellar hemisphere
- C. Cerebellar vermis
- D. Motor relay of the thalamus
- E. Premotor area
- F. Sensory motor cortex
- G. Superior colliculus/pretectal area
- H. Supplementary motor area
- I. Ventral horn

*Insert patient vignette describing a patient with a problem.* The most likely cause is occlusion of which of the following arteries?

- A. Left anterior cerebral
- B. Right anterior cerebral
- C. Left middle cerebral
- D. Right middle cerebral
- E. Left posterior cerebral
- F. Right posterior cerebral
- G. Left lenticulostriate
- H. Right lenticulostriate

*Insert patient vignette describing a patient with a problem.* Which of the following is the most likely diagnosis?

- A. Amyotrophic lateral sclerosis
- B. Huntington's disease
- C. Multiple sclerosis
- D. Parkinson's disease
- E. Poliomyelitis
- F. Polyneuropathy
- G. Retinitis pigmentosa

### Sample Item Set

An unresponsive 58-year-old woman is brought to the emergency department after collapsing at a local shopping mall. Her family reports that she felt well that morning but developed a progressively severe headache. She has had hypertension and atrial fibrillation and is taking an antihypertensive medication and an oral anticoagulant. Her blood pressure is 220/130 mm Hg and she has apnea alternating with hyperpnea. She responds only to noxious stimuli with extensor posturing involving the right arm and leg. Funduscopic examination shows papilledema involving the left optic disc. Pupils are 3.0/7.0 (R/L) with no reaction to light on the left. There is a left gaze preference. There is diffuse hyperreflexia, R > L, and bilateral Babinski signs are present.

1. The dilated, unreactive left pupil is most consistent with injury to which of the following structures on the left?
  - A. Optic nerve
  - B. Optic tract
  - C.\* Oculomotor nerve
  - D. Lateral geniculate nucleus
  - E. Superior colliculus
2. The extensor posturing on the right is most consistent with injury to which of the following areas of the brain on the left?
  - A. Telencephalon
  - B. Diencephalon
  - C.\* Midbrain
  - D. Pons
  - E. Medulla
3. Which of the following best describes her respiratory pattern?
  - A.\* Cheyne-Stokes
  - B. Central neurogenic hyperventilation
  - C. Apneustic
  - D. Ataxic
4. Which of the following herniation syndromes is most consistent with her clinical presentation?
  - A. Cingulate gyrus beneath the falx
  - B.\* Temporal lobe uncus across the tentorium
  - C. Diencephalon through the tentorial notch
  - D. Brain stem through the tentorial notch
  - E. Cerebellar tonsils through the foramen magnum

# Pathology

*Insert description of "action".* Which of the following cellular organelles are most directly involved?

- A. Golgi complex
- B. Lysosomes
- C. Peroxisomes
- D. Mitochondria
- E. Nuclear envelope
- F. Secretory (zymogenic) granules
- G. Rough endoplasmic reticulum
- H. Smooth endoplasmic reticulum

*Insert patient vignette describing a patient with a problem.* Which of the following is the most likely finding on examination of tissue obtained on renal biopsy?

- A. Acute necrotizing vasculitis
- B. Amyloid deposits
- C. Cortical necrosis
- D. Glomerular mesangial IgA
- E. Glomerular peripheral IgG
- F. Granulomas
- G. Interstitial neutrophils
- H. Nodular glomerulosclerosis
- I. Regenerating tubules
- J. Tubular uric acid crystals

*Insert patient vignette describing a patient with a problem.* Which of the following endocrine abnormalities is most likely to be present?

- A. Adrenal cortical adenoma
- B. Adrenal cortical carcinoma
- C. Adrenal hyperplasia
- D. Adrenal necrosis
- E. Adrenal neuroblastoma
- F. Adrenal pheochromocytoma
- G. Pancreatic islet cell adenoma
- H. Parathyroid adenoma
- I. Parathyroid carcinoma
- J. Parathyroid hyperplasia

A 32-year-old man dies 4 days after fracturing his femur in an automobile collision. Examination of his brain at autopsy shows widespread petechiae in the cerebral white matter. Which additional autopsy finding is most likely?

- A. Adult respiratory distress syndrome
- B. Contrecoup injury
- C.\* Fat embolization
- D. Septicemia
- E. Subdural hematoma

Autopsy of a 24-year-old woman shows pleuritis, membranous thickening of glomerular capillary walls, concentric rings of collagen around splenic arterioles, and excrescences on the underside of the mitral valve. Which of the following findings is most likely on evaluation of blood from this woman?

- A.\* Antinuclear antibody
- B. Increased C3 concentration
- C. Lymphocytosis
- D. Monoclonal gammopathy
- E. Positive bacterial culture

A patient with hepatitis B dies 9 days after the onset of symptoms. Which of the following findings is most likely on microscopic examination of his liver?

- A. Diffuse fatty vacuolization with minimal necrosis
- B. Diffuse fibrosis with nodularity
- C. Limited peripheral zonal necrosis
- D.\* Widespread hepatocellular necrosis
- E. Widespread neutrophilic infiltration in lobules

# Pathology Mechanisms

## Template for Mechanisms Items

Item Stem (Patient Vignette): (Describe a patient with a problem)

Lead-In: Which of the following mechanisms is the most likely cause of the patient's findings?

Options: (List of mechanisms)

*Insert patient vignette which describes a patient with a problem.* Which of the following is the most likely cause of the myocardial disease?

- A. Alcohol toxicity
- B. Cardiac amyloidosis
- C. Endomyocardial fibrosis
- D. Hemochromatosis
- E. Hypertrophic cardiomyopathy
- F. Löffler's endocarditis
- G. Postviral myocarditis
- H. Sarcoidosis
- I. South American trypanosomiasis (Chagas' disease)
- J. Vitamin B<sub>1</sub> (thiamine) deficiency

*Insert patient vignette describing a patient with jaundice.* Which of the following is the most likely cause of the jaundice?

- A. Alcoholic hepatic disease
- B. Drug reaction
- C. Dubin-Johnson syndrome
- D. Gilbert's syndrome
- E. Hemolytic jaundice
- F. Intrahepatic ductal atresia
- G. Obstructive jaundice
- H. Primary biliary cirrhosis
- I. Viral hepatitis

*Insert patient vignette describing a patient with a hemostatic abnormality. Which of the following is the most likely cause of the hemostatic abnormality?*

- A. Acute disseminated intravascular coagulation
- B. Factor V (proaccelerin) deficiency
- C. Factor VII (proconvertin) deficiency
- D. Hemophilia A
- E. Hemophilia B
- F. Idiopathic thrombocytopenia purpura
- G. Pulmonary thromboembolism
- H. von Willebrand's disease

*Insert patient vignette describing a patient with a problem. Which of the following is the most likely gastrointestinal lesion?*

- A. Candidal esophagitis
- B. Diverticulitis
- C. Hiatal hernia
- D. Peptic ulcer
- E. Pseudomembranous enterocolitis
- F. Pyloric stenosis
- G. Regional enteritis
- H. Subphrenic abscess
- I. Ulcerative colitis

# Pathology Diagnosis

*Insert patient vignette to describe a patient with a histologic abnormality.* Which of the following is the most likely thyroid disorder?

- A. Chronic autoimmune (Hashimoto's) thyroiditis
- B. Graves' disease
- C. Myxedema after  $^{131}\text{I}$  therapy
- D. Subacute thyroiditis
- E. Well-differentiated thyroid carcinoma

*Insert patient vignette describing a patient with a problem.* Which of the following is the most likely diagnosis?

- A. Acute lymphoblastic leukemia
- B. Acute myelogenous leukemia
- C. AIDS
- D. Chronic lymphocytic leukemia
- E. Ewing's sarcoma





A 38-year-old woman has congestive heart failure, premature ventricular contractions, and repeated episodes of ventricular tachycardia. Her blood pressure is normal. Her heart is markedly enlarged. There are no murmurs; coronary angiography is normal. Which of the following is the most likely diagnosis?

- A. Acute rheumatic fever
- B. Congenital fibroelastosis
- C. Constrictive pericarditis
- D.\* Myocardial infarction
- E. Primary cardiomyopathy

A 74-year-old man has colicky abdominal pain in the left lower quadrant, leukocytosis, and fever. He has not had diarrhea or constipation. Which of the following is the most likely diagnosis?

- A. Carcinoma of the sigmoid colon
- B.\* Diverticulitis
- C. Familial adenomatous polyposis
- D. Ulcerative colitis
- E. Villous adenoma of the upper rectum

# Pathophysiology

## Template for Pathophysiology Items - Additional Findings

Item Stem (Patient Vignette): (Describe a patient with a problem)

Lead-In: Which of the following additional findings is most likely to be present?

Options: (List of findings)

Laboratory tests on an edematous 35-year-old man show a normal serum concentration of complement and an increased serum concentration of cholesterol. Urinalysis show proteinuria (4+), 0-5 erythrocytes/hpf, and several hyaline casts. Which of the following findings is most likely on renal biopsy?

- A. Acute poststreptococcal(proliferative) glomerulonephritis
- B. Membranoproliferative glomerulonephritis
- C.\* Membranous glomerulonephritis
- D. Minimal change disease
- E. Rapidly progressive glomerulonephritis

A 16-year-old boy is undergoing evaluation for jaundice. Laboratory studies show normal hepatic enzyme activities, a negative direct antiglobulin test, increased mean corpuscular hemoglobin concentration, and increased osmotic fragility of erythrocytes. Which of the following types of erythrocyte is most likely to be seen on a peripheral blood smear?

- A. Ovalocyte
- B. Schistocyte
- C.\* Spherocyte
- D. Target cell
- E. Teardrop cell

An asymptomatic 50-year-old woman has hypertension. Urinary excretion of catecholamines is increased. A CT scan shows a suprarenal mass. Which of the following findings is most likely on microscopic examination of the resected mass?

- A. Benign neoplasm of the adrenal cortex
- B.\* Benign neoplasm of the adrenal medulla
- C. Malignant neoplasm of the adrenal cortex
- D. Malignant neoplasm of the adrenal medulla
- E. Diffuse hyperplasia of the adrenal cortex
- F. Diffuse hypoplasia of the adrenal medulla

*Insert patient vignette describing a patient with a problem.* Which of the following is the most likely pulmonary finding at autopsy?

- A. Soft with fluffy consistency, multiple blebs over the surface
- B. Very heavy; frothy and bloody fluid exudes freely from the cut surface
- C. Diffuse fibrotic nodules; whorled appearance replacing normal lung parenchyma
- D. Liver-like consistency in left lower lobe; microscopically contains fibrin and neutrophils in the alveoli
- E. Normal-appearing; large Y-shaped, coiled, laminated clot almost filling both pulmonary arteries

*Insert patient vignette describing a patient with a problem.* Which of the following sets of findings is most likely?

<u>Effective circulating volume</u>	<u>Extracellular fluid volume</u>	<u>Plasma volume</u>	<u>Urine Na<sup>+</sup> excretion</u>
A. Decreased	decreased	decreased	decreased
B. Decreased	increased	decreased	decreased
C. Decreased	increased	increased	decreased
D. Increased	increased	increased	increased

*Insert patient vignette describing a patient with a problem.* Which of the following digestive enzymes or cofactors is most likely to be involved?

- A. Amylase
- B. Chymotrypsin
- C. Colipase
- D. Enterokinase
- E. Lactase
- F. Lipase
- G. Pepsin
- H. Sucrase
- I. Trypsin

*Insert patient vignette describing a patient with a problem.* Which of the following laboratory findings is the most likely?

- |                    |                  |
|--------------------|------------------|
| A. Hypercalcemia   | E. Hypernatremia |
| B. Hypocalcemia    | F. Hyponatremia  |
| C. Hypermagnesemia | G. Hyperkalemia  |
| D. Hypomagnesemia  | H. Hypokalemia   |

*Insert patient vignette describing a patient with a problem.* Which of the following is the most likely thyroid function profile?

- |    | <u>Thyroxine (T<sub>4</sub>)</u> | <u>Triiodothyronine (T<sub>3</sub>)</u> | <u>Resin Uptake</u> | <u>Thyroid-stimulating Hormone</u> | <u>T<sub>3</sub></u> |
|----|----------------------------------|---|---------------------|------------------------------------|----------------------|
| A. | ê                                | ê                                       |                     | ê                                  | ê                    |
| B. | ê                                | normal                                  |                     | é                                  | ê                    |
| C. | ê                                | normal                                  |                     | é                                  | normal               |
| D. | ê                                | Đ é                                     |                     |                                    | Đ                    |



# Pharmacology

## Template for Adverse Effects Pharmacology – "Guess My Drug"

Item Stem (Patient Vignette): (Describe a patient with adverse drug effect)

Lead-In: Which of the following drugs has the patient most likely been taking?

Options: (List of drugs)

A patient in the emergency department cannot remember which "heart drug" he is taking. He says he has "ringing in his ears." His heart rate is greater than 80/min. The ECG shows prolonged PR and QRS intervals. Which of the following drugs has the patient most likely been taking?

- A. Digoxin
- B. Lidocaine
- C. Phenytoin
- D. Propranolol
- E.\* Quinidine

*Insert patient vignette describing a patient with an adverse drug effect.* Which of the following is the most likely to have caused the adverse effect?

- |                         |                   |
|-------------------------|-------------------|
| A. Acetaminophen        | J. Nalidixic acid |
| B. Amiodarone           | K. Nitrofurantoin |
| C. ACE inhibitors       | L. Penicillin     |
| D. Aspirin              | M. Prednisone     |
| E. Atenolol             | N. Procainamide   |
| F. Bleomycin            | O. Propranolol    |
| G. Cytosine arabinoside | P. Sulfasalazine  |
| H. Furosemide           | Q. Tetracycline   |
| I. Metronidazole        | R. Verapamil      |

Drug Y has a volume of distribution ( $V_d$ ) of 75 L in both younger and older adult men. In younger adults, it has a clearance rate of 15L/h, 50% of which is via the liver and 50% via the kidneys. For younger men, the maintenance regimen is 100 mg every 6 hours. Which of the following regimens will produce essentially the same steady-state concentration in an older man, whose creatinine clearance is half that of younger men, but whose hepatic function is unimpaired?

- A. 75 mg every 3 hours
- B.\* 75 mg every 6 hours
- C. 75 mg every 9 hours
- D. 100 mg every 3 hours
- E. 100 mg every 6 hours
- F. 100 mg every 12 hours

A 24-year-old sexually active woman reports increased vaginal discharge. Pelvic examination shows a green frothy discharge. Microscopic examination of a wet mount of the discharge shows motile unicellular organisms 10-30  $\mu$  in length. Which of the following features of this disorder best explains why the infection can be effectively treated with metronidazole?

- A. Facultatively aerobic
- B. Microaerophilic
- C. Strictly aerobic
- D.\* Strictly anaerobic

**Template for Site/Mechanism of Action Items – Pharmacology**

Item Stem (Patient Vignette): (Describe a patient who requires pharmacotherapy)

Lead-In: A drug with which of the following sites/mechanisms of action is most likely to be effective?

Options: (List of mechanisms of action) OR (List of sites of action)

# Pharmacotherapy

## Template for Pharmacotherapy Items

Item Stem (Patient Vignette): (Describe a patient who requires pharmacotherapy)

Lead-In: Administration of which of the following is most appropriate?

Options: (List of drugs)

A 40-year-old black 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. Administration of which of the following is most appropriate?

- |                       |                              |
|-----------------------|------------------------------|
| A. Alpha agonists     | G. Central sympatholytics    |
| B. Alpha blockers     | H. Direct vasodilators       |
| C. ACE inhibitors     | I. Negative inotropic agents |
| D. Beta agonists      | J. Thiazide diuretics        |
| E. Beta blockers      | K. Vasoconstrictors          |
| F. Cardiac glycosides |                              |

*Insert patient vignette describing a patient who requires pharmacotherapy.* Which of the following is the most appropriate initial drug therapy?

- |                |                  |
|----------------|------------------|
| A. Adenosine   | G. Naloxone      |
| B. Aspirin     | H. Prednisone    |
| C. Caffeine    | I. Propranolol   |
| D. Epinephrine | J. Quinidine     |
| E. Insulin     | K. Streptokinase |
| F. Lidocaine   |                  |



*Insert patient vignette describing a patient who requires pharmacotherapy. Which of the following is the most appropriate therapeutic agent?*

- A. Carbamazepine
- B. Dextroamphetamine
- C. Ethosuximide
- D. Haloperidol
- E. L-dopa/carbidopa
- F. Lithium carbonate
- G. Phenobarbital
- H. Primidone
- I. Propranolol
- J. Pyridostigmine

*Insert patient vignette describing a patient who requires pharmacotherapy. Which of the following is the most appropriate therapeutic agent?*

- A. Atropine
- B. Bismuth subsalicylate
- C. Cimetidine
- D. Diphenoxylate
- E. Kaolin
- F. Milk of magnesia
- G. Misoprostol
- H. Omeprazole
- I. Ranitidine
- J. Sucralfate

*Insert patient vignette describing a patient who requires pharmacotherapy. Which of the following is the most appropriate diuretic?*

- A. Carbonic anhydrase inhibitor
- B. High-ceiling or loop
- C. Nonsteroidal potassium-sparing
- D. Osmotic
- E. Steroidal potassium-sparing
- F. Thiazide
- G. Xanthine

# Physiology

An anesthetized patient is being mechanically ventilated. Initial arterial blood gas values are normal. If the ventilation is decreased, which of the following best describes arterial  $\text{PCO}_2$  and pH.

- | Arterial $\text{PCO}_2$ | pH        |
|-------------------------|-----------|
| A. Decrease             | decrease  |
| B. Decrease             | increase  |
| C.* Decrease            | no change |
| D. Increase             | decrease  |
| E. Increase             | increase  |
| F. Increase             | no change |

*Insert patient vignette describing a patient with a problem.* Which of the following substances is most likely involved?

- |                               |                        |
|-------------------------------|------------------------|
| A. ADH (vasopressin)          | E. Bradykinin          |
| B. Aldosterone                | F. Calcitonin          |
| C. Angiotensin                | G. Parathyroid hormone |
| D. Atrial natriuretic peptide | H. Renin               |

A 22-year-old man with a 3-week history of polyuria and polydipsia has had nausea, vomiting, and decreased responsiveness for the past 12 hours. Urinalysis shows 4+ glucose and 4+ ketones. Which of the following sets of arterial blood gas findings is most likely?

- |     | <u>pH</u> | <u><math>\text{PO}_2</math> (mm Hg)</u> | <u><math>\text{PCO}_2</math> (mmHg)</u> | <u><math>\text{HCO}_3</math> (mEq/L)</u> |
|-----|-----------|---|---|--|
| A.  | 7.15      | 98                                      | 33                                      | 11                                       |
| B.* | 7.15      | 98                                      | 24                                      | 8  |
| C.  | 7.30      | 56                                      | 80                                      | 38                                       |
| D.  | 7.40      | 100                                     | 40                                      | 25                                       |
| E.  | 7.50      | 100                                     | 33                                      | 25                                       |
| F.  | 7.50      | 100                                     | 24                                      | 18                                       |
| G.  | 7.50      | 56                                      | 33                                      | 25                                       |

*Insert patient vignette describing a patient with a problem.* Which of the following humoral substances is most likely involved?

- A. Acetylcholine
- B. Adenosine
- C. Aldosterone
- D. Bradykinin
- E. Epinephrine
- F. Norepinephrine
- G. Prostacyclin (PGI<sub>2</sub>)
- H. Prostaglandin E<sub>2</sub>
- I. Prostaglandin F<sub>2</sub>"
- J. Serotonin
- K. Thromboxane A<sub>2</sub>

# Integrative Item Sets

## Template for Integrative Indications Items: Microbiology/Pharmacology

Item Stem (Patient Vignette) (Describe a patient with a problem)

A 40-year-old man with AIDS has a 1-week history of low-grade fever and lethargy. Temperature is 38 C (100.4 F), pulse is 90/min, and blood pressure is 110/70 mm Hg. There is slight resistance to passive flexion of the neck. Laboratory stud-

### Sample Item Set

A 34-year-old woman has had severe watery diarrhea for the past 4 days. Two months ago, she had infectious mononucleosis. She abuses drugs intravenously and is seropositive for HIV. Physical examination shows dehydration and marked muscle weakness.

1. Which of the following laboratory abnormalities is most likely?
  - A. Decreased serum  $K^+$  concentration
  - B. Decreased serum  $Ca^{2+}$  concentration
  - C. Increased serum  $HCO_3^-$  concentration
  - D.\* Increased serum  $Na^+$  concentration
  - E. Increased serum pH
2. In evaluating the cause of the diarrhea, which of the following studies is most appropriate?
  - A. Colonic biopsy to identify *Giardia lamblia*
  - B. Culture of the oral mucosa for *Candida albicans*
  - C. Duodenal biopsy to identify *Entamoeba histolytica*
  - D. Gastric aspirate to identify *Mycobacterium avium-intracellulare*
  - E.\* Stool specimen to identify *Cryptosporidium*
3. Further studies to evaluate her HIV infection show a helper to suppressor T lymphocyte ratio of 0.3. Which of the following actions of HIV best explains this ratio?
  - A. Induction of helper T lymphocyte proliferation
  - B. Induction of suppressor T lymphocyte proliferation
  - C.\* Infection of cells with CD4 receptor
  - D. Infection of macrophages
  - E. Stimulation of leukotriene synthesis

### **Template for Pathology/Pharmacology Items**

Item Stem (Patient Vignette): (Describe a patient with a problem)

Lead-In: Which of the following is the most likely diagnosis?

Options: (List of diagnoses)

Next Lead-In: Administration of which of the following drugs is most appropriate?

Next List of Options: (List of drugs)

### **Sample Item Set**

1. A 62-year-old man with alcohol dependence is admitted to the hospital for transurethral resection of the prostate. The following morning, while being transported to the operating room, he has two generalized seizures within 5 minutes. Neurologic examination shows no focal abnormalities. Which of the following is the most likely diagnosis?
  - A.\* Alcohol withdrawal
  - B. Korsakoff's syndrome
  - C. Partial complex seizure
  - D. Wernicke's encephalopathy
2. The most appropriate management is intravenous administration of which of the following drugs?
  - A. Diazepam
  - B. Haloperidol
  - C. Phenobarbital
  - D. Phenytoin
  - E. Valproate

## Sample Item Set

A 2-year-old boy has had recurrent infections since he was 6 months old. Serum complement concentrations, phagocytic function, and bactericidal activity of neutrophils are normal. A skin test with *Candida* antigens results in 2 cm of induration by 48 hours. Which of the following provides the best explanation for the time of onset of the disease at 6 months of age?

- A. Development of a viral infection
- B. Exposure to rare bacteria
- C. Loss of passive immunity from the mother
- D. A maturation defect in the thymus
- E. Transplacental transfer of IgM antibodies

Which of the following laboratory tests is most likely to be abnormal?

- A. Activity of myeloperoxidase
- B. CD4/CD8 T lymphocyte ratio
- C. Sweat chloride concentration
- D. Fc receptors on macrophages
- E. Serum IgG concentration

## Template for Integrative Physiology/Pharmacology Items

Item Stem (Patient Vignette): (Describe a patient with a problem)



# Sample Item-Writing Option Lists For the Clinical Sciences

## Abdominal Mass

- A. Abdominal abscess
- B. Acute lymphocytic leukemia
- C. Carcinoma of the colon
- D. Carcinoma of the ovary
- E. Choledochal cyst
- F. Constipation
- G. Desmoid tumor
- J. Hernia
- K. Hydatid cyst
- L. Inflammatory bowel disease
- M. Neuroblastoma
- N. Pancreatic pseudocyst
- O. Polycystic kidney
- P. Uterine fibroids

## Abdominal Pain

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

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

## Altered Mental Status

- A. Brain abscess
- B. Cerebrovascular occlusion/  
transient ischemic episode
- C. Dehydration
- D. Drug overdose/drug toxicity
- E. Hepatic encephalopathy
- F. Hypercalcemia
- G. Hyperglycemia
- H. Hyperkalemia
- I. Hypertensive encephalopathy
- J. Hypoglycemia
- K. Hyponatremia
- L. Hypothyroidism
- M. Multi-infarct dementia
- N. Primary degenerative dementia, Alzheimer type
- O. Seizure
- P. Sepsis
- Q. Uremia
- R. Wernicke's encephalopathy

For each patient with altered mental status, select the most likely diagnosis.

## Anemia

- A. Acute myeloblastic leukemia
- B. Aplastic anemia
- C. Chronic lymphocytic leukemia
- D. Drug-induced immune hemolytic anemia
- E. Falciparum malaria
- F. Folate deficiency
- G. Glucose 6-phosphate dehydrogenase deficiency
- H. Hereditary spherocytosis
- I. Iron deficiency anemia
- J. Malignancy metastatic to bone marrow
- K. Microangiopathic hemolytic anemia
- L. Multiple myeloma
- M. Myelofibrosis
- N. Sickle cell disease
- O. Thalassemia minor
- P. Vitamin B<sub>12</sub> (cyanocobalamin) deficiency

For each patient with anemia, select the most likely diagnosis.

## Back Pain

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

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

## Bizarre Behavior

- A. Acidosis
- B. Acute psychosis
- C. Acute subdural hematoma
- D. Alcohol intoxication
- E. Cocaine ingestion/overdose
- F. Delirium tremens
- G. Depression
- H. Hypoglycemia
- I. Hypoxia
- J. LSD ingestion
- K. Malingering
- L. Meningitis
- M. Reye's syndrome
- N. Steroid psychosis
- O. Subarachnoid hemorrhage
- P. Temporal lobe seizure
- Q. Tricyclic antidepressant overdose

For each patient with bizarre behavior, select the most likely diagnosis.

## Breast Lump / Complaint

- A. Breast cyst
- B. Fat necrosis of the breast
- C. Fibroadenoma of the breast
- D. Galactocele
- E. Gynecomastia
- F. Inflammatory carcinoma of the breast
- G. Intraductal papilloma
- H. Lipoma of the breast
- I. Mastodynia
- J. Metastatic carcinoma of the breast
- K. Paget's disease of the breast
- L. Puerperal mastitis
- M. Tietze's syndrome

For each patient with a breast-related problem, select the most likely diagnosis.

## Chest Pain

- A. Angina pectoris
- B. Compression fracture of the spine
- C. Dissecting aortic aneurysm
- D. Esophageal spasm
- E. Esophagitis
- F. Herpes zoster
- G. Hyperventilation
- H. Myocardial infarction
- I. Pectoral muscle pain
- J. Pericarditis
- K. Pneumonia
- L. Pneumothorax
- M. Pulmonary embolism

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

## Diarrhea

- A. Amebiasis
- B. Bacterial gastroenteritis
- C. Carcinoid syndrome
- D. Carcinoma of the colon
- E. Chronic pancreatitis
- F. Crohn's disease
- G. Cryptosporidium infection
- H. Diverticulitis
- I. Dumping syndrome
- J. Gastric lymphoma
- K. Ischemic colitis
- L. Laxative abuse
- M. Pseudomembranous colitis
- N. Short gut syndrome
- O. Sprue
- P. Villous adenoma
- Q. Viral gastroenteritis
- R. Zollinger-Ellison syndrome

For each patient with diarrhea, select the most likely diagnosis.

## **Fatigue**

- A. Acute leukemia
- B. Anemia: chronic disease
- C. Congestive heart failure
- H. Hereditary spherocytosis
- I. Hypothyroidism
- J. Iron deficiency

## Fever in Children

- A. Acute lymphocytic leukemia
- B. Bacterial meningitis
- C. Chickenpox
- D. Drug fever
- E. Gastroenteritis
- F. Juvenile rheumatoid arthritis
- G. Kawasaki disease
- H. Mastoiditis
- I. Measles
- J. Osteomyelitis
- K. Otitis media
- L. Pneumococcal pneumonia
- M. Pyelonephritis
- N. Rheumatic fever
- O. Roseola
- P. Rubella
- Q. Sinusitis
- R. Streptococcal infection

For each patient with fever, select the most likely diagnosis.

## Gastrointestinal Bleeding

- A. Amebiasis
- B. Angiodysplasia of the colon
- C. Clostridium difficile colitis
- D. Carcinoma of the colon
- E. Carcinoma of the esophagus
- F. Carcinoma of the stomach
- G. Coagulopathy
- H. Diverticulitis
- I. Epistaxis
- J. Esophageal varices
- K. Hemorrhoids
- L. Inflammatory bowel disease
- M. Ischemic colitis
- N. Mallory-Weiss tear
- O. Peptic ulcer disease
- P. Reflux esophagitis
- Q. Salmonella infection
- R. Shigella infection

For each patient with gastrointestinal bleeding, select the most likely diagnosis.

## Headache

- A. Aseptic meningitis
- B. Bacterial meningitis
- C. Brain abscess
- D. Brain metastases
- E. Cerebrovascular occlusion
- F. Cluster headache
- G. Drug-induced headache
- H. Dysfunction of temporomandibular joint
- I. Hypertensive crisis
- J. Migraine
- K. Primary brain tumor
- L. Pseudotumor cerebri
- M. Subacute bacterial endocarditis
- N. Subarachnoid hemorrhage
- O. Temporal arteritis
- P. Tension headache
- Q. Trigeminal neuralgia

For each patient with headache, select the most likely diagnosis.

## Leg Pain

- A. Acute arterial insufficiency
- B. Buerger's disease
- C. Causalgia
- D. Cellulitis
- E. Chronic arterial insufficiency
- F. Chronic venous stasis
- G. Dermatomyositis
- H. Diabetic neuropathy
- I. Frostbite
- J. Leriche's syndrome
- K. Lymphedema
- L. Obstruction of the inferior vena cava
- M. Osteomyelitis
- N. Pressure ulcer
- O. Raynaud's disease
- P. Thrombophlebitis
- Q. Varicose veins
- R. Venous insufficiency

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



## **Limb Weakness**

- A. Brachial plexopathy
- B. C1-2 subluxation
- C. C7-T1 subluxation
- D. Encephalopathy
- E. Epidural abscess
- F. Fracture of the cervical spine
- G. Guillain-Barré syndrome
- H. Lead neuropathy
- I. Migraine
- J. Multiple sclerosis
- K. Myasthenia gravis
- L. Osteoarthritis
- M. Poliomyelitis
- N. Postictal state
- O. Ruptured cervical disc ischemia
- P. Stroke/transient cerebrovascular
- Q. Subdural hematoma
- R. Vertebral metastases

## Respiratory Problems

- A. Allergic reaction
- B. Aspiration pneumonia
- C. Bronchiectasis
- D. Carcinoma of the lung
- E. Chronic obstructive pulmonary disease
- F. Congestive heart failure
- G. Cystic fibrosis
- H. Drug side effects
- I. Exercise-induced asthma
- J. Foreign body
- K. Gastroesophageal reflux
- L. Mitral stenosis
- M. Occupational asthma
- N. Pleurisy
- O. Pulmonary embolism
- P. Respiratory syncytial virus
- Q. Sarcoidosis
- R. Tuberculosis

For each patient with respiratory symptoms, select the most likely diagnosis.

## Shock

- A. Addisonian crisis
- B. Anaphylaxis
- C. Atrial fibrillation
- D. Atrioventricular block
- E. Bleeding peptic ulcer
- F. Carcinoid syndrome
- G. Cardiac tamponade
- H. Cardiomyopathy
- I. Dehydration
- J. Fracture of the pelvis
- K. Myocardial infarction
- L. Pulmonary embolus
- M. Ruptured abdominal aortic aneurysm
- N. Ruptured abscess of the appendix
- O. Ruptured ectopic pregnancy
- P. Ruptured ovarian cyst
- Q. Small bowel obstruction

For each patient with shock, select the most likely diagnosis.

## Shortness of Breath

- A. Abscess of the lung
- B. Adult respiratory distress syndrome
- C. Anemia
- D. Asthma
- E. Chronic obstructive pulmonary disease
- F. Congestive heart failure
- G. Diaphragmatic rupture
- H. Foreign body
- I. Hyperventilation
- J. Laryngeal spasm
- K. Mesothelioma
- L. Myasthenia gravis
- M. Pancoast's tumor
- N. Pleural effusion
- O. Pneumothorax
- P. Primary pulmonary hypertension
- Q. Pulmonary embolism
- R. Pulmonary fibrosis

For each patient with shortness of breath, select the most likely diagnosis.

## Urinary Problems

- A. Acute cystitis
- B. Acute urinary retention
- C. Carcinoma of the bladder
- D. Carcinoma of the prostate
- E. Diabetes insipidus
- F. Diabetes mellitus
- G. Hypercalcemia
- H. Interstitial cystitis
- I. Neurogenic bladder
- J. Psychogenic polydipsia
- K. Prostatic hyperplasia
- L. Prostatitis
- M. Pyelonephritis
- N. Renal cell carcinoma
- O. Urethral stricture

For each patient with a urinary problem, select the most likely diagnosis.

## Urinary Symptoms

- A. Acute urinary retention
- B. Bladder fistula
- C. Carcinoma of the bladder
- D. Cystitis
- E. Drug effect
- J. Myoglobinuria
- K. Nephrolithiasis
- L. Pneumaturia
- M. Prostatism
- N. Renal carcinoma

### **Vaginal Discharge/Pruritus**

- A. Allergic vaginitis
- B. Atrophic vaginitis
- C. Bacterial vaginitis
- D. Breakthrough bleeding
- E. Candidal vaginitis
- F. Cervical carcinoma
- G. Condyloma acuminatum
- H. Endometrial cancer
- I. Foreign body
- J. Gonococcal cervicitis
- K. Normal menses
- L. Placenta previa
- M. Threatened abortion
- N. Trichomoniasis
- O. Vesicovaginal fistula
- P. Vulvar carcinoma

For each patient with vaginal discharge or pruritus, select the most likely diagnosis.

### **Wheezing**

- A. Angioedema
- B. Asthma
- C. Bronchiectasis
- D. Carcinoid syndrome
- E. Chronic obstructive pulmonary disease
- F. Congestive heart failure
- G. Cystic fibrosis
- H. Endobronchial polyp
- I. Epiglottitis
- J. Esophageal reflux
- K. Foreign body
- L. Goiter
- M. Laryngeal dyskinesia
- N. Pneumonia
- O. Pulmonary edema
- P. Pulmonary embolism
- Q. Respiratory syncytial virus infection
- R. Viral croup

For each patient with wheezing, select the most likely diagnosis.