

**A GUIDE TO PREPARING
MULTIPLE-CHOICE ITEMS**



P · E · S

© 2005 by the Professional Examination Service, Inc. All rights reserved. No part of this document may be reproduced in any form without the permission in writing of Professional Examination Service, Inc. For information, contact: Professional Examination Service, Inc., 475 Riverside Drive, New York, NY 10115.

INTRODUCTION

The most commonly used type of test item in licensure and certification examinations is the conventional multiple-choice item. By following the instructions in this manual, test developers and representatives of professional associations will increase their proficiency in writing and reviewing this type of item. On the surface, writing multiple-choice items may appear relatively simple, given familiarity with the subject matter, but experienced test authors find that a great deal of skill, patience, and creativity is required.

In the instructional sections that follow, general knowledge is used as the content of the examples, in order to facilitate understanding of the general principles. The skills needed for writing quality items are similar across content disciplines.

Like any other craft, multiple-choice item writing requires a significant amount of painstaking practice and appropriate feedback. It is essential to try out your items on others. This process will help to uncover ambiguities in wording and unintended violations of item writing principles. Feedback will greatly enhance item quality and writer/reviewer productivity.

[NOTE: Professional Examination Service (PES) encourages feedback from the users of this document, so that relevant, desirable changes may be incorporated into the next revision.]

PARTS OF THE MULTIPLE-CHOICE ITEM

All multiple-choice items consist of two basic parts: the **stem** (or premise) and the **choices** (or options).

The **stem** is the introductory statement or question that elicits the correct (desired) answer.

The **choices** are suggestions for completing the statement or answering the question presented by the stem. Only one choice is the correct answer. In each example that follows, four choices (1 correct and 3 incorrect) are presented.

[NOTE: In each sample item in this manual, the correct answer appears as the first choice. Please enter the correct answer as the first choice on the item construction form.]

BASIC SAMPLE ITEMS

STEM Which of the following scientists was responsible for the development of the first effective polio vaccine?

CHOICES

1. Jonas Salk
2. Louis Pasteur
3. Edward Jenner
4. Robert Koch

In the example above, the correct choice is called the **correct answer** and the three incorrect choices are called **distractors**. A major factor in multiple-choice item writing is the ability to create high-quality distractors that effectively discriminate between candidates who have mastered the material being assessed and those who have not.

An item stem can be written in the form of an incomplete statement as well as in question form. The item above could also be presented in this way:

STEM The first effective polio vaccine was developed by:

CHOICES

1. Jonas Salk.
2. Louis Pasteur.
3. Edward Jenner.
4. Robert Koch.

[Note that each choice in this item is followed by a period because each choice serves to complete a sentence.]

It is wise to draft an item first as a direct question; revise it to an incomplete sentence if this form offers smoother, less repetitive wording.

Whichever format is used, remember that text **clarity** is the objective. Clarity can often be checked by covering the choices and reading the stem alone, to determine whether an informed candidate would be able to answer the question before reading the choices.

CORRECT- VS. BEST-ANSWER ITEMS

It is important for item writers to keep in mind the distinction between items in which **only one** choice is correct, and items in which more than one choice may be technically correct but only one is clearly the **best** answer.

Correct-Answer Item

The capital of Kentucky is:

1. Frankfort.
2. Lexington.
3. Louisville.
4. Ashland.

Best-Answer Item

The major goal of raising the legal drinking age is to:

1. reduce driving fatalities and injuries.
2. reduce liquor consumption.
3. discourage the habit of drinking by teenagers.
4. reduce roadside litter.

In writing best-answer items, word the stem in a way that clearly indicates that only **one** choice is preferred. In the example above, specifying “**The major goal**” helps to achieve this objective.

REFINING MULTIPLE-CHOICE ITEMS

The following principles are essential in developing high-quality multiple-choice items that do not confuse or penalize the candidate, nor reward the candidate by inadvertently providing clues that make the correct answer obvious. Typically, the correct answer is easier to create than the distractors. Remember that poorly crafted distractors increase the ability of the candidate to guess the correct answer. For a well-constructed four-choice multiple-choice item, the chance of guessing the correct answer is 1 in 4 (25%). One poor distractor increases the chance to 1 in 3 (33%); with two poor distractors, it is 1 in 2 (50%); with three poor distractors, the answer is obvious and a correct guess becomes 100% certain. **While you are acquiring item-writing experience, refer to the principles below frequently. Apply them when evaluating items written by others.**

1. Use appropriate language.
 - a. Employ standard American English grammar, usage, and spelling throughout the stem and the choices. (Program-specific instructions will be supplied for examinations administered in other countries, if necessary.) Each choice must smoothly follow or connect with the stem, both logically and grammatically.
 - b. Always take into account the relative reading ability (i.e., school grade level) and the degree of knowledge and experience of the candidate population (i.e., whether the examination is designed for entry-level or specialist candidates). Most PES examinations are for entry-level candidates, i.e., those who just meet the eligibility requirements, set by the relevant professional association, for taking the examination.
 - c. Context, setting, content, and language should be equally appropriate and familiar to all properly trained candidates. Avoid vocabulary and constructions that might favorably or unfavorably bias the performance of specific segments of the candidate population, including members of minorities; for example, avoid wording that is peculiar to one geographic region or is unnecessarily technical.

Poor

A biological explanation for sex-related differences in human behavior would most likely be supported by research on:

1. fetally androgenized females.
2. cultural conditioning.
3. pseudohermaphrodites.
4. neuroendocrinology.

Better

A biological explanation for sex-related differences in human behavior would most likely be supported by research on:

1. sex-hormone-level disorders during the fetal stage.
2. cultural stereotypes of males and females.
3. people who have both male and female physical characteristics.
4. comparisons of male and female athletic performance.

d. Avoid vocabulary and descriptions related to ethnicity, culture, disability, religious or political affiliation, age, gender, etc., that may be considered offensive. (See below concerning gender-neutral language.) Do not reinforce stereotypes. Aside from issues of sensitivity and fairness, the performance of a candidate who experiences an emotional reaction may be impaired.

e. Use a “person first” presentation; e.g., “a patient who has diabetes,” instead of “a diabetic patient.”

2. Present the question in the item stem as clearly and succinctly as possible.

a. Leave out nonessential information, **but** make sure to include **all** of the information needed to identify the correct answer. Test, don't teach; extraneous factual or descriptive information can delay the candidate's progress through the examination and may tip off the correct answer to another item. Feedback from others is especially useful in making these determinations.

Poor

In 1850, Adolphe Chattin, then professor of pharmacy in Paris, believed that goiter resulted from an inadequate amount of iodine in the diet.

The thyroid is part of which body system?

1. Endocrine
2. Nervous
3. Musculoskeletal
4. Cardiovascular

Better

The thyroid is part of which body system?

1. Endocrine
2. Nervous
3. Musculoskeletal
4. Cardiovascular

b. Include in the stem all words that would otherwise have to be repeated in each of the choices.

Poor

The process whereby amniotic fluid is removed from a pregnant woman to test for possible birth defects:

1. is known as amniocentesis.
2. is known as cesarean section.
3. is known as embryonic analysis.
4. is known as fetal catheterization.

Better

The process whereby amniotic fluid is removed from a pregnant woman to test for possible birth defects is known as:

1. amniocentesis.
2. cesarean section.
3. embryonic analysis.
4. fetal catheterization.

c. If the stem requires the narration of a sequence of events, use the present tense: "A happens, then B happens, then still later C happens."

d. At times, a stem may contain stimulus material such as photographs, charts, diagrams, or drawings, to help to condense and clarify information.

3. The stem should pose **only one question or problem**. Its presentation should be so clear, specific, and informative that, even before reading the choices, the candidate can anticipate the type of answer desired. The candidate can be easily confused or misdirected by an ambiguous stem for which several choices could be considered correct.

Poor

Diabetes is a:

1. disorder associated with the body's absorption of glucose.
2. digestive disorder that affects adults.
3. potentially fatal disorder if not detected early.
4. disease that can result in blindness.

Better

Diabetes is a chronic disease characterized by:

1. the inability of the body to absorb glucose.
2. a genetic abnormality resulting in high blood sugar levels.
3. an excessive ingestion of sweet foods during infancy.
4. an inability to digest amino acids.

Poor

Alcoholic beverage consumption is believed to be a:

1. tradition in Western cultures.
2. vice by those with certain religious beliefs.
3. means to achieve desirable states of relaxation.
4. major cause of family conflict and disorganization.

Better

Studies on alcoholic beverage consumption in the United States during the past 10 years suggest what statistical trend?

1. Increased binge drinking by young adults
2. Decreased consumption by adolescents
3. Increased abstinence among all age groups
4. Increased heart disease risk among those over age 65

[Note that in the examples above, making the stem more specific significantly clarifies the identity of the desired answer.]

4. State the stem positively; a negative stem tends to confuse the candidate and therefore reduce the effectiveness of the item in assessing subject matter knowledge and understanding. Double-negative constructions can be particularly confusing.

Poor

Which of the following authors is **not** a major American novelist?

1. George Orwell
2. Saul Bellow
3. John Steinbeck
4. Ernest Hemingway

Better

Which of the following authors is a major British novelist?

1. George Orwell
2. Saul Bellow
3. John Steinbeck
4. Ernest Hemingway

The following type of presentation is sometimes useful:

Also Better

All of the following authors are major American novelists **except**:

1. George Orwell.
2. Saul Bellow.
3. John Steinbeck.
4. Ernest Hemingway.

[Note that it is acceptable to ask the candidate for negative information when it is **critical** to know what **not** to do, for example, to prevent harmful drug interactions (e.g., “Which of the following additional medications would be **contraindicated** for a patient who is currently taking Drug A?”).]

5. Construct gender-neutral items; reword to eliminate any gender references that are unnecessary and could convey discrimination or stereotyping.

Write in the third person. Repeat the noun referring to a person (e.g., “the therapist”) in order to avoid using “he” and “she” and their gender-related adjectives. If the resulting phrasing is still extremely awkward, “he/she” may be used in a hypothetical context. Also, try changing the number of persons from singular to plural.

[Note that it is inappropriate to use “he/she” or “they” and related adjectives to refer to one person.]

In the rare instance when a proper name must be supplied for an individual, use a gender-neutral one (e.g., Pat).

Poor

If an elementary school teacher has a student who has been stung by a bee, she should do which of the following things first?

1. Determine if he is allergic to bee stings.
2. Remove the stinger from her with tweezers.
3. Attempt to kill the bee.
4. Bandage the affected area.

Better

Which of the following is the first thing an elementary school teacher should do when a student has been stung by a bee?

1. Determine if the student is allergic to bee stings.
2. Remove the stinger with tweezers.
3. Attempt to kill the bee.
4. Bandage the affected area.

It is often possible to present and request information in an item without mentioning a specific person; for example, if mention of a school setting is unnecessary, the following type of presentation could be used:

Also Better

Which of the following is the first thing to do when an individual has been stung by a bee?

1. Determine if the individual is allergic to bee stings.
2. Remove the stinger with tweezers.
3. Attempt to kill the bee.
4. Bandage the affected area.

[Use of the passive voice is sometimes helpful, but should not be very frequent.]

6. In general, base items on facts and on currently accepted principles, practices, and theories of the profession.

a. The correct answer must clearly stand out as the one that experts in the field would recognize as the **best** choice. When professional doubt or controversy exists about the correct answer, the item should not be submitted. If a candidate becomes confused, the item could be open to challenge, and it might have to be omitted before final scoring of the examination.

Poor

The primary cause of the Vietnam war was:

1. Vietnamese nationalism.
2. French imperialism.
3. Chinese expansionism.
4. American imperialism.

Better

Which of the following factors most likely sustained the military campaign of the Vietnamese insurgents in the Vietnam war?

1. Nationalism
2. Chinese expansionism
3. Soviet militarism
4. American expansionism

b. If an item refers to a professional evaluation or judgment (not merely an opinion), be sure that the stem identifies an appropriate authority or source (see also “6c.” below, about citation). This specification is especially important in dealing with controversial subject matter and with decisions regarding ethics or interpersonal relationships, because otherwise the correctness of the “best” answer could be open to challenge.

Poor

The best way to discipline a child is to:

1. reward good behavior.
2. use physical punishment for serious offenses.
3. explain why an undesirable behavior is unacceptable.
4. ignore misbehavior.

Better

According to the theories of behavioral psychology, the best way to discipline a child is to:

1. reward good behavior.
2. use physical punishment for serious offenses.
3. explain why an undesirable behavior is unacceptable.
4. ignore misbehavior.

c. The correct answer to an item must also be supported by current authoritative references. Supply full identification of at least one (ideally, the most relevant) authoritative, current, easily accessible, professionally respected resource in the “Reference” section of the item construction form. Include author/editor name, title, edition/volume, publisher, year, and page. An Internet citation should include the appropriate URL, not simply mention “the so-and-so Website.”

d. The reference should not simply mention a phrase or clause that appears in the item, but should provide a description or discussion that adequately covers the topic and agrees with the item’s correct answer. The reference may discuss a concept, but the actual item can ask for a definition of the concept or, better, for an application of the concept to a job-related situation. Note that verbatim repetition of the wording in the reference is inappropriate (see “7b.” below).

7. Remove clues that make it easy for the candidate to detect the correct answer without having the knowledge or skill that the item is intended to measure.

a. All choices should be parallel in concept, grammar, and structure.

Poor

An American city noted for its production of automobiles is:

1. Detroit.
2. Ohio.
3. Indiana.
4. Philadelphia.

Better

An American city noted for its production of automobiles is:

1. Detroit.
2. Cincinnati.
3. Indianapolis.
4. Philadelphia.

Poor

Serious communicable diseases that can cause developmental disability to the offspring of a pregnant mother are:

1. rubella and chickenpox.
2. cancer.
3. rheumatism and arthritis.
4. sinus infection.

Better

A serious communicable disease that can cause developmental disability to the offspring of a pregnant mother is:

1. rubella.
2. cancer.
3. rheumatism.
4. sinus infection.

Write correct answers that have approximately the same length, detail, and technical content as the distractors. If this is not possible, try to create two relatively short and two relatively long choices, two relatively simple and two relatively complex choices, or two relatively technical and two relatively nontechnical choices. Do not include one distractor that opposes or contradicts the correct answer unless the two remaining distractors also oppose or contradict each other.

b. A correct answer taken from a textbook or other major source should be rephrased rather than quoted directly. The wording in reference materials tends to be more technical than the wording in made-up distractors and may be easily identified, or even remembered, by a candidate who has little or no content knowledge.

Poor

What is anorexia?

1. A psychophysiological condition of post-pubertal adolescence in which caloric intake is reduced
2. A prolonged period of nausea
3. An adolescent growth disorder
4. A disorder of the pituitary gland

Better

What is anorexia?

1. Lack or loss of the appetite for food
2. A prolonged period of nausea
3. An adolescent growth disorder
4. A disorder of the pituitary gland

c. Do not use the same descriptive or key words in the stem and the correct answer but not in the distractors.

- d. Do not make the correct answer clear and concise and the distractors vague and ambiguous.
- e. Do not use relative terms in the stem and the correct answer (e.g., usually, most) but absolute terms in the distractors (e.g., always, every).

8. Ideally, distractors should present unsafe practices, commonly held misconceptions, or true but irrelevant information; in any case, they should always be **plausible**. Avoid using distractors that even the most uninformed candidate would recognize as being incorrect, but **do not attempt to trick** candidates into choosing incorrect choices.

[Note that the use of humorous or absurd distractors is not appropriate in standardized test items.]

Poor

The person who developed the psychoanalytic theory of human behavior is:

1. Sigmund Freud.
2. C.G. Jung.
3. Dr. Joyce Brothers.
4. Dr. Phil.

Better

The person who developed the psychoanalytic theory of human behavior is:

1. Sigmund Freud.
2. C.G. Jung.
3. B.F. Skinner.
4. Karen Horney.

9. Choices (especially correct answers) should not overlap in content or completely include the text of other choices. If a distractor is a subset of the correct answer, a candidate might select the distractor and still technically be partially correct. Be especially careful about overlap when the choices include numbers or numerical ranges.

Poor

The right to vote in the United States is granted to individuals of what age?

1. 18
2. 16
3. 17
4. 19

Better

The right to vote in the United States is granted to individuals **beginning** at what age?

1. 18
2. 16
3. 17
4. 19

[Note that in the example on the left, both number 1 and number 4 could be considered correct.]

10. **“All of the above”** and **“none of the above”** are inappropriate as choices. It is logically difficult to claim that one of the other choices in the item is either correct or incorrect. Appropriately choosing **“none of the above”** as the correct answer does not indicate that the candidate knows correct information, only that he or she recognized wrong choices. **“Two of the above”** and **“More than one of the above”** are also unacceptable. Note also that all of these wordings contradict the requirement that each item must have one and only one **“best”** answer.

11. Avoid using phrases like “What would you do?” or “What do you believe?” With this type of question, the candidate could always defend his or her selected choice as correct, since each person has a unique frame of reference. Use the prescriptive terms “should” or “ought to” or “must” instead. Ask the candidate to identify what is the best or most appropriate or most commonly accepted practice for a professional in the field.

12. Items should be worded so that they will not be out of date when they appear on an examination.

Poor

In the next-to-last presidential election, a vice presidential candidate who was defeated was:

1. Richard Cheney.
2. Joseph Lieberman.
3. Ralph Nader.
4. Bill Bradley.

Better

In the presidential election of 2000, a vice-presidential candidate who was defeated was:

1. Richard Cheney.
2. Joseph Lieberman.
3. Ralph Nader.
4. Bill Bradley.

13. Items should assess knowledge of meaningful facts, concepts, and professional practices, not trivial information.

Poor

The assassination of President John F. Kennedy occurred on what date?

1. November 22, 1963
2. December 4, 1961
3. February 18, 1961
4. June 7, 1959

Better

President John F. Kennedy was assassinated:

1. before the passage of the major civil rights legislation of the 1960's.
2. before the Bay of Pigs invasion of Cuba.
3. after the Vietnam war ended.
4. before the construction of the Berlin Wall.

[Note that the item on the right assesses the sequence of historical events rather than a single date.]

Asking the candidate to identify the “least important”/“least common”/“least appropriate” choice does not determine whether the candidate knows what is “most important.”

Do not make up names of people, laws, diseases, assessment instruments, etc., just to meet the quota of distractors. (This practice is acceptable when an item is asking what the letters in an acronym stand for, but this type of information is often trivial.)

14. “Fill-in-the-blanks” construction, with gaps in the middle of the stem that are completed by the choices, is unacceptable.

15. Some programs require items that assess a candidate's response to a job-related situation. It can be hard to create distractors that produce an item with the right degree of difficulty. First, try writing a question asking the best way a professional would respond to the given situation. Next, prepare a list of choices with a wide range of acceptability among experienced professionals in the appropriate occupational group. Finally, present the question and list of choices to several professionals from that occupational group and have them select what they believe to be the best answer and three reasonable distractors. This is a time-consuming practice, but it is essential for writing items of this kind. It is also important for the stem in this type of item to provide only essential information.

16. Write some items for **each of the cognitive levels** (see Appendix 2). In addition to items that ask for knowledge/recall of basic information (Level 1), write some that will assess the higher-level cognitive skills such as application/interpretation of knowledge (Level 2) and problem solving/evaluation (Level 3). (Some programs use a two-level system, with knowledge/recall as Level 1 and all of the higher functions combined into Level 2.) The higher-level items are needed to assess important components of candidate competence.

Knowledge

Compared to boys, how much earlier or later is the average start of puberty for girls?

1. 2 years earlier for girls
2. 3 years later for girls
3. 1 year earlier for girls
4. 2 years later for girls

Application of Knowledge

A chaperone at a seventh-grade dance would generally observe which of the following situations?

1. The girls are taller than the boys.
2. The boys are taller than the girls.
3. The girls and the boys are of approximately equal height.
4. Most of the boys have completed puberty.

Problem Solving

The director of a school musical wants the male and female chorus members to be of approximately equal height. This goal could best be achieved by auditioning:

1. seventh-grade girls and eighth-grade boys.
2. seventh-grade boys and ninth-grade girls.
3. ninth-grade boys and ninth-grade girls.
4. eighth-grade girls and seventh-grade boys.

17. Occasionally, it is useful to base a group of items (a **series**) on one information set (called a **scenario**). A scenario can present a context or relate the items to a commonly encountered situation.

The items in a series must be independent of each other, i.e., to determine the correct answer to a series item, the candidate must obtain information only from the scenario and from the item's own stem. In addition, if any item is dropped permanently from the series or is not used on a particular examination, the items that do appear must still make sense without the missing one.

If the series represents a logical or temporal sequence (for example, diagnosis, treatment, and follow-up), identify each item as "1 of 2," "3 of 4," and so forth, on the item construction form, so that the sequence can be restored if the construction forms become separated during review and editing.

If this format is used, always make sure that the information presented in one item does not give away the correct answer of a related item.

[Heading in examination booklet:]

**QUESTIONS 1 THROUGH 3 REFER
TO THE FOLLOWING INFORMATION:**

[Scenario:]

On a biology examination, students obtained the following scores: 90, 88, 85, 76, and 72.

[Items:]

1. What is the average or mean score for this group of students?

1. 82.2
2. 85.0
3. 79.5
4. 84.8

SEE INFORMATION PRECEDING QUESTION 1

2. If the average score is 82.2, what is the median score?

1. 85
2. 88
3. 76
4. 90

[Note that Item 2 is poor because it provides the answer to item 1.]

SEE INFORMATION PRECEDING QUESTION 1

3. What is the range of the scores?

1. 18 points
2. 90 points
3. 72 points
4. 4 points

USEFUL RESOURCES ON MULTIPLE CHOICE ITEM WRITING:

Case S M and Swanson D B (2003) Constructing Written Test Questions for the Basic and Clinical Sciences, 3rd Edition, Revised. Philadelphia, PA: National Board of Medical Examiners.

[Available in PDF format only, requiring Adobe Acrobat Reader 4.0 or higher.

Go to: <http://www.nbme.org/about/itemwriting.asp>

Accessed June 27, 2005.]

Haladyna T M (2004) Developing and Validating Multiple-Choice Test Items, 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

CHECKLIST FOR MULTIPLE-CHOICE ITEMS

ITEM CONSTRUCTION FORM

- Is each item typed (double-spaced) or written clearly on an individual item construction form?
- Does the item contain the appropriate number of choices? (See program-specific construction form for the number.)
- Is the correct answer presented as the first choice and is it marked with an asterisk?
- Are the areas of content outline (see Appendix 1, Rubrics) and any program-specific codes clearly indicated? (Program-specific codes will be explained separately at workshops.)
- Is the cognitive level designation (if required) clearly indicated?
- Has at least one current, authoritative resource been fully cited in the “Reference” section?

STYLE AND CONTENT

Entire Item

- Have standard American English grammar, usage, and spelling been used throughout the item (except when indicated otherwise for tests administered in other countries)?
- Is the language appropriate for the general reading level of the candidate population?
- Is the language appropriate for the degree of professional expertise/experience of the candidate population?
- Is the item worded clearly and concisely?
- Is the item free from potentially offensive, derogatory, stereotyping, and performance-biasing language?
- Has all excessively technical language or jargon been eliminated?
- In programs that have specific style guidelines, have the guidelines been consulted?
- Does the item seek meaningful information, not trivia?
- Are the stem and the answer based on established fact or, alternatively, on current, generally accepted theory, principles, or practices of the profession?
- Is the item free from inappropriate humor?
- If the item includes measurements, have the units been indicated?
- Are series items independent of each other?
- Do some of the items assess application and others problem solving, in addition to items that assess knowledge? (See Appendix 2, Cognitive Levels)

Stem

- Does the stem clearly address only one problem or content area?
- Has the stem been stated positively rather than negatively?
- Do the wording and structure of the stem avoid tipping off the correct answer or too easily ruling out any of the distractors?

Choices

- Does the item have one and only one correct or best answer?
- When the correct answer is based on fact or theory from a specific authority or source, has this been cited in the stem and supported by at least one solid reference?
- Does each choice match the stem logically and grammatically, and follow or join it smoothly?
- Are all of the choices parallel in structure, length, degree of technicality, and amount of detail?
- Are all of the distractors likely to be plausible to the less-competent candidate but definitely wrong?
- Are all of the choices independent of each other?
- Has the inclusion of “all of the above” and “none of the above” choices been avoided?

(Reviewers will also consider the points presented in Appendix 3, Reviewer Validity Ratings.)

APPENDIX 1 RUBRICS

A rubric is a numeric designation that identifies the domain, the task, and the knowledge on which an item is based. Before a program's rubrics are created, a role delineation or practice analysis identifies the knowledge, skills, tasks, responsibilities, and ethical considerations involved in proper performance of the profession. This information is coded in a system of even-numbered strings of digits (the rubrics) spanning the range of practice and identifying specific combinations of each of these components.

For example, in psychopharmacology, which uses a 6-digit rubric, the rubric 020324 is built from the following three components:

1. Domain 02 (the first pair of digits) refers to the overall function "interpret, generate, disseminate knowledge in pharmacotherapy." Within this domain are various task and knowledge statements, including:
2. Task 03 (the second pair of digits) refers to "prepare and disseminate drug information."
3. Knowledge 24 (the last pair of digits) refers to "knowledge of media systems...for reporting...drug research."

Some programs add a secondary rubric, also consisting of an even number of digits, to further refine coverage of the profession.

[Item writers and reviewers will receive program-specific instructions on accurate assignment of rubrics to items.]

APPENDIX 2 COGNITIVE LEVELS

To determine the cognitive level of an examination item, read it and ask yourself what a candidate has to do mentally in order to pick the correct answer. There are three (3) basic cognitive levels:

Level 1: Knowledge/Recall

The candidate needs only to define a term, recall a fact, or recognize the information in the correct answer that belongs with the information in the stem. This level basically involves previously learned information.

Level 2: Application/Interpretation

The candidate needs to apply a principle or theory to answer the question, calculate an equation, organize information in a new way, or analyze complex material. This level requires understanding of information and the ability to relate it to other information.

Level 3: Problem-Solving/Evaluation

The candidate needs to formulate a procedure or plan to solve a problem, synthesize information into a comprehensive whole, or compare sets of information to arrive at the correct answer. This level requires reorganization of information into new information, or the formulation of judgments, for use in a new situation.

Each level has some conceptual overlap with the next higher one.

Of the programs that indicate levels in the items, some differentiate only two levels, the original Level 1 and a combination of original Levels 2 and 3.

Whether or not a program indicates level designations in items, it is important for the bank and for each examination to include items that represent all of these functions.

APPENDIX 3 REVIEWER VALIDITY RATINGS

In addition to the topics covered above in “Refining Multiple-Choice Items” and “Checklist for Multiple-Choice Items,” item reviewers must evaluate the information (knowledge, skills, performance) in submitted items with regard to the following questions:

1. Would a candidate who lacked this information be likely to cause harm to a client/patient, the professional organization, or the general public?
2. How frequently is this information required/used in practice?
3. Is it possible to practice the profession capably and responsibly without having this information?
4. In addition to the vocabulary in the item, do the concepts and procedures apply everywhere in the area in which the candidate would be practicing the profession?
5. Does the item measure what it is intended to measure?

Program-specific validity rating scales and reviewer checklists will be available for reviewers. The set of scales on the following page is an example.

SAMPLE ITEM VALIDATION RATING SCALES

A. Is the knowledge or skill measured by this item **relevant** to entry-level performance as a certified professional?

1. No
2. Not sure
- 3.** Yes

(Proceed to Scales B and C if Response 3 on Scale A was selected.)

B. How **important** is the knowledge or skill tested by this item to the assessment of entry-level performance as a certified professional?

1. Not important
2. Of little importance
- 3.* Moderately important
- 4.* Very important
- 5.** Of greatest importance

C. What **degree of harm** would result from performance errors caused by lack of the knowledge or skill measured by this item?

1. No harm
2. Minimal harm
- 3.* Moderate harm
- 4.** Considerable harm

(“Harm” can be understood to mean a detrimental effect; categories include, but are not limited to, physiological, psychological, or economic.)

* indicates an “Adequate” rating

** indicates an “Optimal” rating