Item Writing Guide

Prepared by

PSI/AMP
Psychometrics Division

www.goAMP.com

TABLE OF CONTENTS

INTRODUCTION

This guide is designed to familiarize you with the overall process of writing an item that can be used for examinations developed by Applied Measurement Professionals, Inc., a PSI business (PSI/AMP). It is not a comprehensive discussion of item writing or examination development; many books are available for the interested reader. This guide will include a discussion of: 1) item types, 2) item cognitive complexity, and 3) item writing suggestions.

While unique terminology may be introduced throughout this guide, an understanding of the following key terms is important to establish uniformity:

Item - the entire question, including the stem and options. In multiple-choice testing it is customary to speak of test "items" rather than questions, since items may be presented in the form of statements rather than questions.

Stem - the statement, question, chart or graph portion of an item. The stem of the item should clearly present the central problem or idea.

Options - all possible answers to the item, including the *distractors* (the incorrect answers to the item), and the *key* (the one correct, best answer to the item).

The cornerstone of any examination program is the individual item. Each item must be linked to a required area of work or practice. Tasks required for work or practice should be identified on the basis of a job analysis and are the justification for the content outline used for the examination. All items should be linked directly to the content outline or other job analysis information. As discussed further in the following sections, all items will be of a four-option, multiple-choice format. Other formats may have a place in some testing programs, but for many reasons, these will not be used for the certification program for which you have been asked to write items.

ITEM TYPES

PREFERRED ITEM TYPE – One Best Response, Positively Worded Stem

This is the traditional and most frequently used type of multiple-choice item, and all items will fit this format, or one of the variations described in this section. The one best response item type consists of a stem, followed by a series of possible answers or completions, called options.

Stem: In 1985, the World Series was won by the

Options:
A. Kansas City Royals. KEY
B. Atlanta Braves. DISTRACTORS (B, C, and D)
C. New York Yankees.
D. St. Louis Cardinals.

In this type of item, the instructions to the examinees emphasize the importance of selecting the "one best response" from among those offered. In answering these items, the examinee is instructed to look for the **best** or **most appropriate** choice and to discard others that may appear plausible, but are less applicable. The two primary formats of the one best response item type are the *Direct Question* and the *Incomplete Statement*, as shown in the following examples.

Direct Question

What is the capitol of the United States?

A. Washington, DC
B. New York City, NY
C. Philadelphia, PA
D. Richmond, VA

Incomplete Statement

Carl Rogers developed a counseling style referred to as

A. Gestalt Therapy.
B. Reality Therapy.
C. Client-Centered Therapy.
D. Rational-Emotive Therapy.

Use either a direct question or an incomplete statement as the item stem, whichever seems more appropriate for effective presentation of the item. If it seems to make little or no difference which item type is used, select the style that you can handle most effectively. Those who have not had experience in writing multiple-choice items may find that they will initially produce fewer technically weak items when they write direct questions rather than the incomplete statement approach. The direct question may induce the item writer to produce more specific options. These items can be rewritten to fit one form or the other. For example, the stem of last item presented above could be rewritten as: *Which of the following counseling styles did Carl Rogers develop?*

You may wish to try rewording a stem as both an incomplete statement and a direct question, then complete the sentence or respond to the item to determine which format you would prefer. Each time you generate an item, consider these alternative formats and the use of different terminology, then attempt to take the perspective of the person who will be asked to respond to the item. While opinions do differ, the version that you feel is clearer and easily understood is probably the best. A special case of the one best response format involves items requiring *calculations*, which often includes *pictorial* or t*abular* information.

ALTERNATE ITEM TYPES – Use with Caution

Situational Set

This type of item can be constructed to measure the examinee's ability to identify, resolve, and manage problems, and to help simulate a realistic situation. Examinees are given a short scenario or an item stem with a collection of facts or data. Then they are presented with multiple-choice items regarding recognition and management of the problem. Each item should be able to stand alone in reference to the situation. The reading and evaluation of interpretive material in the item stem consumes a relatively large amount of testing time, but its inclusion in an examination is justified by the number of items that are based on it – usually three to five. Do not provide a series of items that give or suggest the response to another item in the set. Likewise, do not prepare a series of items in which a correct response to one item is necessary to determine the correct response to another.

The scenario in a situational set should always meet the following criteria:

- It should be professionally accurate.
- It should focus on some central theme.
- It should provide information that is relevant to the associated items.
- It should not include information unnecessary to answer the items which follow.

Situational Set No. 1

A snow removal service provides the following data for the month of January:

Week	No. of Jobs	Weekly Revenue
1	25	$10,800
2	20	$8,000
3	45	$45,700
4	19	$9,500

What was the total revenue for the month of January?

A. $18,500
B. $74,000
C. $296,000
D. $888,000

Which week was the MOST productive in terms of amount of money per job?

A. 1
B. 2
C. 3
D. 4

If number of jobs is indicative of snowfall, which week had the largest snowfall?

A. 1
B. 2
C. 3
D. 4

Situational Set No. 2

A study designed to compare the length of hospital stay with total charges produces the following data:

Length of Stay	Charges
2 days	$1800
4 days	$5000
5 days	$5700
7 days	$7500
8 days	$9000

What is the mean length of stay?

A. 2.6 days
B. 5.0 days
C. 5.2 days
D. 8.0 days

What is the mean charge per day?

A. $5000
B. $5400
C. $5800
D. $6200

Which of the following would best estimate the correlation coefficient between length of stay and charges?

A. 0.00
B. + 0.25
C. + 0.50
D. + 0.99

These examples show the appropriate format for a situational set: each item relates to the scenario, and each item is independent from the other in terms of determining the correct answer. However, what should be noted is that this type of item does not have to involve calculation. It is acceptable to include only text in the items.

The format alternatives discussed in this section are the only ones recommended. As mentioned previously, other formats or item types may have a place in examinations, but they are not generally acceptable for certification examinations developed by PSI/AMP.

ITEM COGNITIVE COMPLEXITY

PSI/AMP uses a three-level classification system to identify the level of thinking required to respond to an item. The three levels are recall, application, and analysis. Identifying the cognitive level of an item will require you to take the perspective of the examinee, that is, to consider the type of thinking required of a typical examinee responding to the item. PSI/AMP certification examinations include items that require more than just recall.

Level 1 - Recall

Recall items primarily test the recognition or recall of isolated information. Such items require predominantly an effort of memory. They include the recall of specific facts, generalizations, concepts, principles, processes, procedures, or theories. To simplify, such an item will ordinarily be asking: "What is X?"

Which state produces the most dairy products?

A. Iowa
B. Ohio
C. Wisconsin
D. California

Which of the following is the formula used to calculate the volume of a room?

A. $2 \times \pi \times r$
B. $\pi \times r^2$
C. L x W x H
D. $(H \times L)^2$

Level 2 - Application

Application items primarily test interpretation or application of limited data. Such items require more than simple recall, which means that they should include some problem-solving. These items require translation into another form of specific verbal, tabular, or graphic data, and recognition of the elements and relationships among such data. Items at this level will ordinarily be asking: "Knowing X to be true, what would you expect to be true about Y?"

A subject in a drug trial was instructed to take 1 tablet of the test article t.i.d. and received 80 tablets 3 weeks ago. After 21 full days, the subject returns for a scheduled visit. How many tablets must the subject return to be considered 100% compliant?

A. 17
B. 26
C. 38
D. 59

From the following employee data, calculate the number of FTEs during a one week (40 hour) pay period.

1 technician with 20 hours
2 technicians with 60 hours each
4 technicians with 50 hours each
8 technicians with 40 hours each

A. 14.5
B. 15.0
C. 16.5
D. 18.0

Level 3 - Analysis

Analysis items primarily test the evaluation of data, problem solving, or the fitting together of a variety of elements into a meaningful whole. Items at this level will ordinarily require examinees to make judgments concerning the effectiveness, appropriateness, or best course of action for a particular situation. Many steps may be required in the thought process of the examinee.

Which of the following performance measures would be used to evaluate the quality of care provided by anesthesiologists?

A. percent of patients developing post-spinal headaches requiring blood patch treatment
B. percent of surgical patients who expire and do not receive an autopsy
C. percent of patients on steroid therapy who develop a renal crisis postoperatively
D. percent of cases in which the pre-operative diagnosis is inconsistent with the post-operative diagnosis

A credentialing examination shows less reliability than desired. One-half of the items have p values greater than .85. Fifteen-percent of the items have negative point-biserial correlations. The psychometrician should recommend that the committee write

A. longer, more difficult to read stems.
B. easier items, focusing on basic facts.
C. clearly wrong, yet attractive distractors.
D. stems requiring complex computations.

While the difficulty of an item is sometimes related to its cognitive level, it must be emphasized that it is the thought process required that determines an item's cognitive level. To write items that assess a particular cognitive level, the essential question to ask is: "What do I expect the examinee to do in order to select the correct response?" Your response to this question is sometimes called your evaluative objective. If you expect the examinee to identify, recall, or recognize, you will generally be writing Level 1 items. If you want the examinee to classify, explain, or differentiate, you are likely writing Level 2 items. If you expect the examinee to formulate, evaluate, or judge, Level 3 items should result.

ITEM WRITING SUGGESTIONS

The Whole Item

1. *EXPOSE ITEMS TO EXPERT REVIEW.* "Review" in this suggestion refers to item content more than it refers to adequacy with respect to grammar or spelling. A formal review of the items should be conducted, and should include:
   - suggestions for better wording
   - verification of the correct answer
   - suggestions for better distractors
   - an elimination of any bias
   - appraisal of an item's significance or relevance to the profession.

2. *FOLLOW THE STANDARD RULES OF GRAMMAR*.
   - Each option should begin with a capital letter and end with a period or other terminal punctuation mark if it is a complete sentence.
   - If the stem of the item is a question and the options are not complete sentences, the options should not use capitalization or terminal punctuation.
   - The period should be omitted with numerical options to avoid confusion with decimal points.
   - When the stem is an incomplete sentence, each option should begin with a lower case letter. No period is used at the end of an incomplete sentence when used as the stem.

3. *AVOID IRRELEVANT SOURCES OF DIFFICULTY.* Just as it is possible to incorporate clues to a correct response inadvertently, it is possible to place obstacles unintentionally. Frequently, problems in mathematics are answered incorrectly by examinees who have reasoned correctly, but slipped in their computations. To measure an understanding of the process used in a calculation, simple numbers (whole numbers) are recommended. Similarly, examinees should not miss an item solely because of language or vocabulary difficulties. All items must be written so that an appropriate reading level is required. You should use appropriate professional terminology, but do not use longer, more complicated words when shorter, less complex words will suffice. Overall, the difficulty of the "average item" should be such approximately 75% of the examinees get it correct. Most item writers tend to underestimate the difficulty of the items they produce.

4. *USE AN EFFICIENT FORMAT.* The options should be listed on separate lines, under one another, like the examples in this guide. This makes the options easy to read and compare. The use of letters in front of the options is preferable to using numbers, since this avoids possible confusion when numerical options are used in an item.

5. *ELIMINATE IRRELEVANT CLUES.* Irrelevant clues may make the item easier as a whole or may even change the basis upon which the item discriminates. "Testwise" examinees who normally would not be able to select the correct option will notice the clue and respond correctly on the basis of it.
   - Similarity of wording in both the stem and the correct answer is one of the more obvious clues. Key words in the stem may be unintentionally repeated verbatim in the correct answer, a synonym may be used, or the words may simply sound or look alike.
   - The phrasing of the correct answer is likely to give it away. Even the most poorly prepared examinees are able to recognize a familiar phrase or "buzzword."
   - When the answer is qualified by modifiers which are typically associated with true statements (e.g., sometimes, may, usually), it is more likely to be chosen.
   - Including absolute terms in the distractors enables examinees to eliminate them because such terms are commonly associated with false statements (e.g., always, never, all, none, only).
   - Including two options that have the same meaning makes it possible to eliminate them as potential answers. If two options have the same meaning and only one answer is to be selected, it is fairly obvious that both of them must be incorrect.

6. *AVOID RARE, EXOTIC, AND TEXTBOOK CASES.* Exceptional cases or examples are indeed just that - *exceptional,* and should be avoided since they do not occur frequently enough to warrant their assessment. Also, these types of cases may be too location or regionally specific and not applicable on a national or company wide basis. Textbook cases encourage rote memorization rather than comprehension, and examinees without a solid understanding of the content of their profession may get the item correct.

The Stem

1. *USE CLEAR AND SIMPLE LANGUAGE.* The production of good test items is one of the most exacting tasks in the field of creative writing. Few other words are read with such critical attention to implied and expressed meaning as those used in test items. The problem of ambiguity in objective test items is particularly acute because each test item is usually an isolated unit. Unlike other reading material in which extensive content and context helps to clarify the meaning of a particular phrase, a test item must be explicitly clear in and of itself.

2. *AVOID UNESSENTIAL SPECIFICITY.* Design your items to test knowledge that may be applied in a variety of specific situations. The superior value of general knowledge over specific knowledge should be reflected in tests whenever possible.

3. *AVOID COMPLEX SENTENCE STRUCTURE.* Unless it is essential for the purpose of the item, sentence structure should be as simple as possible. Complex sentences should be broken up into two or more separate sentences. Qualifying phrases should be

placed near the terms they qualify. The important elements should generally appear early in the statement of the item, with qualifications and explanations following.

4. *PRESENT A SINGLE, CLEARLY FORMULATED PROBLEM.* The task set forth in the stem of the item should be so clear that it is understood without reading the options. In fact, a good check on the clarity and completeness of a multiple-choice stem is to cover the options and determine whether it could be answered. Avoid an "undirected stem" in which the examinee must read all options to know what is being asked in the item.

5. *AVOID THE USE OF NEGATIVE WORDING.* State the stem of the item in positive form whenever possible. A positively-phrased item tends to measure more important learning outcomes than a negatively stated item. This item type generally displays poorer psychometric characteristics because candidates may become confused.

6. *USE of WHAT vs. WHICH in the STEM.* The word "What" implies an absolute answer. If the correct answer is a fact, use the word "What." If the answer requires judgment, or if other answers not listed in the options could be just as good as a possible answer, use the word "Which." The word "Which" is used to limit the realm of choice to the listed options.

7. *AVOID ITEMS RELATING TO DEFINITIONS.* A correct response to a definition simply indicates that the examinee can simply recall something learned in class or training session. It does not mean the examinee can use that information in a job-related setting.

8. *PUT AS MUCH OF THE WORDING AS POSSIBLE INTO THE STEM.* Avoid repeating the same material in each of the options. By moving all common content to the stem, it is usually possible to clarify the problem and reduce the time required to read the options. If necessary reword the entire item to avoid repetition in the options.

9. *AVOID EXCESSIVE "WINDOW DRESSING."* The item should contain only material relevant to its solution, unless selection of relevant material is part of the problem.

10. *INCLUDE ALL QUALIFICATIONS NEEDED TO SELECT THE RIGHT ANSWER.* Item writers do not always state the qualifications that exist in their own minds about a topic. They forget that examinees may need to have these qualifications specifically stated.

Options/Distractors

1. *SELECT AND FORMULATE THE DISTRACTORS WITH CARE.* The distractors are as important as your statement of the problem in the stem. The difficulty of an item depends largely on the distractors. The finer the distinctions that must be made to select the correct answer from the options, the more difficult the item.

2. *MAKE CERTAIN THAT THE KEY IS CORRECT AND CLEARLY BEST.* There should be ONLY ONE correct answer, and it should be unquestionably correct. The intended answer should be the one that experts would agree is clearly the best, in which case, it may be necessary to include "which of the following" in the stem to allow for equally satisfactory answers which have not been included in the item.

3. *MAKE THE DISTRACTORS PLAUSIBLE TO THE UNINFORMED OR MISINFORMED.* The distractors in a multiple-choice item should be so appealing to examinees who lack the knowledge called for by the item that they select one of the distractors in preference to the correct answer. The art of constructing good multiple-choice items depends heavily on the development of effective distractors. Incorrectness is not the sole criterion for the selection of a distractor. The following are a number of things that can be done to increase the plausibility and attractiveness of distractors.
   - Use the common misconceptions of examinees as distractors.
   - State the options in the language of the examinee.
   - Use good-sounding words (e.g., accurate, important, etc.) in the distractors, as well as in the correct answer.
   - Make the distractors similar to the correct answer in both length and complexity of wording.
   - Use extraneous clues in the distractors, such as stereotyped phrasing, scientific-sounding answers, and verbal associations with the stem of the item.
   - Make the options similar but avoid fine discriminations which are not practically significant.
   - Avoid using options which are opposites of each other. Each option should be plausible; opposites are inconsistent with that idea, and examinees can eliminate them with limited information.

4. *ARRANGE THE OPTIONS IN A LOGICAL ORDER.* When the options consist of numbers, they should ordinarily be arranged in ascending order, to make it easier for examinees to locate their choice and mark the answer accurately.

5. *DO NOT USE THE OPTIONS "ALL OF THE ABOVE" OR "NONE OF THE ABOVE."* The inclusion of "all of the above" as an option makes it possible to answer the item on the basis of partial information, and the chances of guessing the correct answer are increased. Another difficulty with this option is that some examinees, recognizing that the first choice is correct, will select it without reading the remaining options. "None of the above" is a poor distractor because in some cases it can be argued that it is the best answer.

6. *MAKE THE OPTIONS INDEPENDENT.* Sometimes a subset of two or three of the options may cover the entire range of possibilities, so that one of them must necessarily be correct. Sometimes one option may include one or more of the other options, so that all of the options in that subset must necessarily be false. In these cases related options help examinees eliminate distractors without relying on content knowledge. When possible, avoid options that are mutually exclusive of each other.

7. *MAKE ALL OPTIONS GRAMMATICALLY CONSISTENT WITH THE STEM AND PARALLEL IN FORM.* Inexperienced item writers often phrase the correct option very carefully so that it is grammatically consistent with the stem, and then pay much less careful attention to the distractors. Make sure that you read *each* option with the stem, and ensure that the options are consistent, parallel, and properly stated.

Summary of Item Writing Suggestions

A. The Whole Item

1. Expose items to expert review.
2. Follow the standard rules of grammar.
3. Avoid irrelevant sources of difficulty.
4. Use an efficient format.
5. Eliminate irrelevant clues.
6. Avoid rare, exotic, and textbook cases.

B. The Stem

1. Use clear and simple language.
2. Avoid unessential specificity.
3. Avoid difficult/complex vocabulary.
4. Present a single, clearly formulated problem. The answer should usually be clear without reading the options.
5. Avoid negative wording.
6. Take care with “what” vs. “which” in the stem.
7. Avoid items relating to definitions.
8. Put as much of the wording as possible into the stem.
9. Avoid excessive “window dressing.”
10. Include all qualifications needed to select the right answer.

C. Options/Distractors

1. Select and formulate the distractors with care.
2. Make certain the key is correct and clearly the best.
3. Make the distractors plausible to the uninformed or misinformed.
4. Arrange the options in a logical order.
5. Do not use “all of the above” or “none of the above.”
6. Make the options independent.
7. Make all options grammatically consistent with the stem, similar in length, and parallel in form.

AMP, a psi business
18000 W. 105th St. • Olathe, Kansas 66061-7543
+1 913 895 4600 • Fax +1 913 895 4650