



Council on Certification

ITEM DEVELOPERS GUIDE

This guide has been prepared to assist individuals in writing test items, particularly for use in credentialing types of examinations. It presents the reasons for using objective tests and, specifically, multiple-choice items (questions). It also describes the characteristics of good items and suggests approaches to developing them.

WHY USE OBJECTIVE TESTS?

An objective examination is one that can be scored on the basis of a predetermined scoring key. Especially when a large group of candidates is to be tested, as in a credentialing program, a machine-scored, objective examination has numerous advantages.

- It is efficient and convenient to administer, since all candidates receive the same questions, the same answer choices, the same type of answer document, and the same directions.
- It eliminates the influence of extraneous variables, such as candidates' handwriting, spelling, and grammar that could interfere with the objective and valid appraisal of the abilities the questions were designed to assess.
- It permits rapid and reliable scoring and reporting of results.
- It facilitates the collection of statistical data about the examination.

WHY USE ONLY ONE TYPE OF TEST ITEM?

Measurement professionals recommend using as small a number of different item types as possible within a single examination, and preferably only one. They base their recommendation on the following premises:

- Administration of the examination is simplified when only one type of item is used. This is especially important when there is a time limit for completing the examination.
- Each different type of item in the examination involves the uses of testing time and space for directions to candidates to explain how to respond to the different item types. From a psychometric standpoint, the time and space could be used more effectively to improve the reliability and validity of the measurement by asking more questions of the one type.
- When candidates must change to a different item type, some may fail to follow directions, leading to errors in measurement.
- Although the inclusion of a variety of item types might be thought to add interest to the examination process, it is equally possible to derive interest from the content and quality of the items, rather than from their form, and a variety of item types may become a source of distraction rather than of interest.

WHY USE MULTIPLE-CHOICE ITEMS?

There are four principal types of objective test items: completion, matching, true-false, and multiple-choice. If only one type of objective test item is to be used, the flexible and adaptable multiple-choice item is the logical choice. A creative item writer can devise multiple-choice items that can be used to measure any of the knowledges, skills, and abilities that can be measured with paper-and-pencil tests. Content that can be tested with a completion item can just as easily be tested with a multiple-choice question by giving a candidate four choices from which to complete the statement rather than having them write out an answer. Matching items can also be adapted into multiple-choice items, and a multiple-choice item can be produced from materials that might have been appropriate for use in four true-false questions, using one true statement as the correct option and three false ones as distracters. (Distracters are sometimes referred to as foils, incorrect options, or wrong answers.)

Multiple-choice items are suitable for use in evaluating the attainment of all types of learning objectives, from the simplest recall of knowledge through comprehension, interpretation, application, analysis, and synthesis, up to and including evaluation. Except for creativity and writing ability, most educational objectives, including understanding and problem-solving, can be assessed by carefully developed multiple-choice items. The usefulness of multiple-choice items is limited principally by the skill of the item writer.

WRITING ITEMS

The task of the item writer is the most difficult and the most critical in the entire process of test development.

Without good items, it is impossible to have a good examination. The item writer must be familiar with the content to be covered by the examination, choose topics for the items that are to be written, present each question and its right answer clearly and succinctly, and devise plausible but incorrect options to serve as distracters.

HOW SHOULD THE ITEM WRITER BEGIN?

Before starting to write any items, the item writer should become familiar with the outline of content for the examination.

The outline defines the content that the examination is to cover, and the emphasis that is to be given to each part of the outline. The item writer must also consider what is expected of the minimally competent candidate who will take the examination in order to determine the appropriate level of difficulty for the items.

Ideas for items may come from many different sources. The item writer's own experience in the work situation is likely to suggest problems and situations around which items can be written. Keeping in mind that the examination will be administered as part of a procedure for the promotion of the public in terms of safety and quality of service, the item writer can often recall functions, roles, concepts, and abilities that are essential components of knowledge for the entry level or minimally competent candidate. In situations involving safety problems, items can be written to determine whether candidates know what should be done. Other sources of inspiration might include situations in which a less-than-competent person could cause harm, or where an improper sequence or lack of knowledge could lead to an undesirable outcome.

Reading materials, both texts and references, can also suggest topics for items, but it is important to keep in mind that the best items rarely measure rote memory; rather, they attempt to evaluate a candidate's ability to apply principles and demonstrate insights and understanding. Sentences taken from printed materials have not been known to yield good test items.

Having selected a topic on which to write an item, the item writer should consider whether the topic involves knowledge or the application of knowledge that is important in relation to those who will take the examination. The item writer should also make sure that there is general agreement about the topic among knowledgeable people in the subject-matter area, and that there are some misconceptions about the topic. Once that is done, the item writer is ready to start drafting the item.

WHAT ARE THE COMPONENTS OF A MULTIPLE-CHOICE ITEM?

Each multiple-choice item is composed of a stem followed by four OPTIONS

-- one CORRECT OPTION and three DISTRACTERS.

EXAMPLE: STEM Who invented the light bulb?

CORRECT OPTION * 1. Thomas Alva Edison

DISTRACTER 2. Benjamin Franklin

DISTRACTER 3. Alexander Graham Bell

DISTRACTER 4. Leonardo da Vinci

The stem of a multiple-choice item may be presented in question form, as in the previous example, or as an incomplete statement, as shown in the example below.

EXAMPLE: STEM The light bulb was invented by

CORRECT OPTION * 1. Thomas Alva Edison.

DISTRACTER 2. Benjamin Franklin.

DISTRACTER 3. Alexander Graham Bell.

DISTRACTER 4. Leonardo da Vinci.

It is sometimes easier for beginners to produce good items if they use the question form, since each of the options must then be an answer to the question asked in the stem. Certain types of content may also lend themselves more readily to one form than the other. Other than those two considerations, there is no psychometric reason for choosing between the question form and the incomplete statement.

WHAT ARE THE CHARACTERISTICS OF A GOOD ITEM STEM?

▷ ***A good item stem presents a problem situation clearly.***

Knowledgeable candidates should be able to look at the stem and produce the desired response before referring to the option. Consider the following examples:

Poor: An elevated white blood cell count is:

- *1. a sign of infection.
2. dangerous.
3. normal.
4. any count over 20,000.

Better: An elevated white blood cell count is most often symptomatic of:

- *1. infection.
2. diabetes.
3. high blood pressure.
4. respiratory insufficiency.

In the poor item, the stem does not pose a problem; it fails to provide candidates with a frame of reference for answering the question. In the better item, the candidates know that they are looking for something of which an elevated white blood cell count is a symptom. This suggests that the correct option will be in the form of a disease or a pathological state.

▷ ***A good item stem is stated positively rather than negatively.***

Poor: Which of the following is NOT a city on California?

- *1. Boise.
2. Sacramento.
3. San Francisco.
4. Los Angeles.

Better: In what state is the city of Boise located?

- *1. Idaho.
2. Iowa.
3. New Jersey.
4. California.

At times, a negative stem is unavoidable, as in a safety related question where it is important for candidates to know what **NOT** to do in a certain situation. In those cases, the negative term (e.g. **NOT**, **EXCEPT**, **LEAST**, **CONTRAINDICATED**, etc.) should be capitalized and presented in bold face or underlined to call attention to it.

▷ ***A good item stem avoids the use of the pronouns "it", "he", "she", and "you".***

In the case of "it", "he", and "she", the candidate may be uncertain about the referent. Questions that ask "What would you do?" can have no single right answer, since individual candidates could justify their individual responses.

▷ ***A good item stem is presented as simply as possible, including only information that is necessary to understand the problem.***

Poor: Mr. Watson is a 73-year-old widower who has suffered from headaches for several years. His wife died four years ago, as the result of a stroke. They had two children, both now grown and living in other parts of the country. Mr. Watson believes that his headaches may be due to high blood pressure, but tests to date have not confirmed his suspicions. In an effort to establish the cause of Mr. Watson's headaches, what is the first procedure that should be initiated?

Better: What should be the first procedure to be initiated to diagnose the probable cause of chronic headaches in an otherwise healthy 73-year-old man?

▷ **A good item stem includes all words that would have to be repeated in each option.**

Poor: Ice forms on water when

- *1. the temperature falls below 32 degrees F at sea level.
- 2. the temperature falls below 24 degrees F at sea level.
- 3. the temperature falls below 12 degrees F at sea level.
- 4. the temperature falls below 0 degrees F at sea level.

Better: At sea level, ice forms on water when the temperature falls below

- *1. 32 degrees F.
- 2. 24 degrees F.
- 3. 12 degrees F.
- 4. 0 degrees F.

▷ **A good item stem specifies the authority or standard upon which the correct option is based, if the item calls for a judgment.**

Poor: the diet of Americans provides vitamins and minerals in amounts that are

- *1. Adequate for normal nutrition.
- 2. Inadequate for normal nutrition.
- 3. In excess of normal requirements.
- 4. Variable in relation to individual requirements.

Better: According to the American Medical Association, the diet of the average American provides vitamins and minerals in amounts that are

- *1. Adequate for normal nutrition.
- 2. Inadequate for normal nutrition.
- 3. In excess of normal requirements.
- 4. Variable in relation to individual requirements.

▷ **A good item stem poses a problem to which the correct answer is not likely to change over time.**

Poor: Who was President of the United States last year?

- *1. Bill Clinton
- 2. George W. Bush
- 3. Gerald Ford.
- 4. Ronald Reagan.

Better: Who was President of the United States in 2009?

- *1. Barack Obama.
- 2. Bill Clinton.
- 3. Gerald Ford.
- 4. Ronald Reagan.

▷ **A good item stem focuses on important learning objectives and avoids testing trivia.**

Testing for knowledge of facts and principles is not trivial, since such evaluation constitutes an appropriate objective.

Poor: The initials for NRA stand for:

- *1. National Recovery Act
- 2. National Rifle Association
- 3. Northeast Registrars Association
- 4. Nebraska Registrars Associations

Better: What legislation, passed by Congress during the Great Depression, was later found to be unconstitutional?

- *1. National Recovery Act
- 2. National Rifle Association
- 3. Northeast Registrars Association
- 4. Nebraska Registrars Associations

The poor item illustrates more than one defect. The item is testing at a trivial level, and any of the four options would be correct. The better item requires candidates to apply more knowledge in selecting the correct option.

WHAT ARE THE CHARACTERISTICS OF GOOD OPTIONS?

- Correct options are unquestionably correct, and distracters are unquestionably wrong.
- All four options are grammatically related to the stem. If the stem asks a question, each of the four options provides a plausible answer to the question. If the stem is an incomplete statement, each option serves to complete the statement.
- The four options are homogeneous in terms of structure as well as content. Difficulty is most likely to occur in achieving this objective when the item stem has not presented a clearly defined question or problem.
- The correct option is similar in length to the distracters. If the item contains distracters that are short and imprecise with a correct option that is long and fully qualified, candidates will quickly recognize and reject the distracters.
- If the correct option includes one or more key words that appear in the stem, the distracters also contain those key words.
- Absolute terms have been avoided. In poorly constructed test items, options containing "all", "none", "always" and "never" are likely to be found in distracters, while less definite terms such as "generally" and "often" are likely to be used in correct options. Candidates will quickly recognize the absolute terms in the distracters and reject those distracters.
- Options are mutually exclusive. If two options have the same meaning, and only one answer is to be selected, candidates will realize that both options must be incorrect and recognize them as distracters.
- The intended correct answer includes no incorrect information which might force candidates to choose between ignoring the incorrect information and selecting some other option.
- "None of the above" has been avoided, if at all possible. If it has been used, then the following additional rules have also been observed.

⊙ The correct option is completely and precisely correct.

⊙ "None of the above" has been used in several items, and it has sometimes been the correct option, but not always.

⊙ "None of the above" has not been used as a fourth option when there are only three logical options, such as "more than", "less than", and "equal to". (In such instances, it is best to discard or revise the item.)

▷ "All of the above" has not been used as an option. The reason for this rule is that if each of the options is a correct response, than any of them would have to be considered as correct in scoring them. Since the examination directions specify that there is a single correct answer to each item, the use of "All of the above" would violate that specification.

HOW CAN AN ITEM WRITER FEEL CONFIDENT THAT AN ITEM IS USABLE?

▷ Follow the foregoing instructions.

▷ After items have been drafted, let them rest for at least one week before reviewing them for final approval. Then review each item singly.

▷ Read the entire item from beginning to end, watching to see if it includes any extraneous or irrelevant material, or omits any essential information. Avoid colloquialisms, expressions, or abbreviations that may derive from regional differences and give some candidates an advantage over others.

▷ Review the stem to be sure it poses a problem or asks a question that establishes a frame of reference for the four options.

▷ Review the options to be sure that all are of similar form. If the stem leads the candidate to expect the name of a disease as the correct option, all four options should be disease names. If the correct option is a quantity, the three distracters should also be quantities.

● Review each option in relation to the stem to be sure each is grammatically and logically related to the stem, and each is plausible, but only one is correct. If the stem is an incomplete statement, each option should complete the statement, and each should agree in terms of singular vs. plural, verb tenses, etc.

▷ If the stem asks "What is ..." be certain that the correct option is the only possible answer. Otherwise candidates could argue the existence of another answer than those provided. If there are other possible correct responses, of which the item includes only one, consider phrasing the stem in terms of "Which of the following ..." While it is generally preferable to avoid "Which of the following..." because it requires candidates to read the options to know the focus of the item, that approach does acknowledge that although there might be other acceptable responses, the candidate is to consider the question in relation to the options presented.

▷ Once again be sure that one, and only one, of the four options is correct and the other three are unquestionably incorrect. Particular care must be taken with items that ask for "the MOST common" or "the LEAST common", etc.

For this reason, those words should be underlined or set in boldface so candidates will seek the BEST response to the item, not merely an adequate response. The item writer must confirm that the correct option is, in fact, the BEST, MOST, LEAST, etc.

▷ If the item writer finds an item has been drafted in which two or three of the options are correct, there may be ways to salvage some or all of the content.

- Combine the two or three correct responses into one option, and try to develop new options containing multiple components to replace the distracters.
- If one of the correct options is better than the others in terms of precision or terminology, discard the others and try to replace them or leave their places blank for colleagues to suggest replacement.
- While negative questions have certain undesirable characteristics, some questions can best be presented in that way. The item writer may choose to ask "Which of the following is NOT ..." Such questions should only be asked when it is important for candidates to know that something is NOT correct.

▷ Once again, check the correct option to be sure it is not conspicuously different from the three distracters, in terms of length, clarity, grammar, and substance.

▷ For final confirmation of the quality of a set of items, many item writers ask a colleague to review their items and attempt to answer them. In this way, defects that may have eluded the creator of the items may be brought to light so they can be corrected before the items are submitted.

WHAT FORMAT SHOULD BE USED IN SUBMITTING ITEMS?

- | |
|--|
| <ul style="list-style-type: none">• Send items to: NCRA's Council on Certification by email – ctrexam@ncra-usa.org.• Indicate the section of the content outline being tested by the item.• Number the options "1.", "2.", "3.", and "4.", placing the numbers one below the other, and starting each option on a new line.• Put the correct option in the first place as option 1., and either circle the option number or place an asterisk before it to confirm that it is the correct option.• Write the source (references + page #) of your item. |
|--|