



**NOT MEASUREMENT
SENSITIVE**

**DOE-HDBK-1204-97
January 1997**

**Supersedes
DOE-STD-1009-92
July 1992**

DOE HANDBOOK

GUIDE TO GOOD PRACTICES FOR THE DEVELOPMENT OF TEST ITEMS



**U.S. Department of Energy
Washington, D.C. 20585**

FSC 6910

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

This document has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from the Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831; (423) 576-8401.

Available to the public from the U.S. Department of Commerce, Technology Administration, National Technical Information Service, Springfield, VA 22161; (703) 487-4650.

Order No. DE97000786

FOREWORD

1. This Department of Energy (DOE) Handbook is approved for use by all DOE Components and their contractors. The Handbook incorporates editorial changes to DOE-STD-1009-92, *Guide to Good Practices for the Development of Test Items*, and supersedes DOE-STD-1009-92. Technical content of this Handbook has not changed from the original technical standard. Changes are primarily editorial improvements, redesignation of the standard to a Handbook, and format changes to conform with current Technical Standards Program procedures.

2. This technical standard, used in conjunction with the DOE *Guide to Good Practices for the Design, Development, and Implementation of Examinations*, can be used as an aid to provide a framework for the systematic development of valid and reliable examinations. Detailed examples of short answer, multiple-choice, and essay test-item formats are included to aid in developing effective test items. Users are not obligated to adopt all parts of this Handbook; rather, they can selectively use the information to establish or improve facility training programs as applicable.

3. Beneficial comments (recommendations, additions, deletions) and any pertinent data that may improve this document should be sent to the Office of Nuclear Safety Policy and Standards (EH-31), U.S. Department of Energy, Washington, DC 20585, by letter or by using the self-addressed Document Improvement Proposal (DOE F 1300.3) appearing at the end of this document.

4. DOE technical standards, such as this Handbook, do not establish requirements. However, all or part of the provisions in a technical standard can become requirements under the following circumstances:
 - (1) They are explicitly stated to be requirements in a DOE requirements document; or
 - (2) the organization makes a commitment to meet a technical standard in a contract or in a plan or program required by a DOE requirements document.

INTENTIONALLY BLANK

TABLE OF CONTENTS

1. INTRODUCTION 1

2. GENERAL DISCUSSION 3

3. DEVELOPMENT OF WRITTEN TEST ITEMS 7

4. SHORT-ANSWER TEST ITEMS 11

5. SHORT-ANSWER TEST ITEM WRITING PROCEDURE 13

6. MULTIPLE-CHOICE TEST ITEMS 17

 6.1 Styles of Multiple-Choice Test Items 19

 6.2 Examples of Multiple-Choice Test Items 23

7. MULTIPLE-CHOICE TEST ITEM WRITING PROCEDURE 25

8. MATCHING TEST ITEM WRITING PROCEDURE 35

9. ESSAY-FORMAT TEST ITEMS 37

 9.1 Styles of Essay-Format Test Items 37

 9.2 Examples of Essay-Format Test Items 38

10. ESSAY-FORMAT TEST ITEM WRITING PROCEDURE 41

11. TEST ITEM ANSWERS AND POINT VALUE 45

 11.1 Cognitive Level vs. Point Values 45

 11.2 Scoring Essay-Format Test Items 47

12. USE OF HANDOUTS 49

13. OPEN-REFERENCE TEST ITEMS 51

APPENDIX A

TEST ITEM DEVELOPMENT FORM A-1

APPENDIX B

ACTION VERBS B-1

APPENDIX C

TEST ITEM EXAMPLES C-1
C.1 Short-Answer Test Item Examples C-3
C.2 Matching Test Item Examples C-7
C.3 Multiple-Choice Test Item Examples C-10
C.4 Essay Test Item Examples C-12

APPENDIX D

EXAM TEST ITEM REVIEW CHECKLIST D-1
D.1 General Guidance D-3
D.2 Point Values D-5
D.3 Short-Answer Test Items D-5
D.4 Multiple-Choice Test Items D-6
D.5 True or False (Multiple-Choice) Test Items D-8
D.6 Matching Test Items D-8
D.7 Essay-Format Test Items D-9
D.8 Drawing or Labeling Test Items D-9

TABLES

Table 1. Tips for Developing the stem of a multiple-choice test item 21

1. INTRODUCTION

While the methodology used in developing test items can vary significantly, to ensure quality examinations, test items should be developed systematically. Test design and development is discussed in the DOE *Guide to Good Practices for Design, Development, and Implementation of Examinations*. This guide is intended to be a supplement by providing more detailed guidance on the development of specific test items.

This guide addresses the development of written examination test items primarily. However, many of the concepts also apply to oral examinations, both in the classroom and on the job. This guide is intended to be used as guidance for the classroom and laboratory instructor or curriculum developer responsible for the construction of individual test items.

This document focuses on written test items, but includes information relative to open-reference (open book) examination test items, as well. These test items have been categorized as short-answer, multiple-choice, or essay. Each test item format is described, examples are provided, and a procedure for development is included. The appendices provide examples for writing test items, a test item development form, and examples of various test item formats.

INTENTIONALLY BLANK

2. GENERAL DISCUSSION

Test items consist of two components: the content (what is asked) and the format (the way it is asked). The quality of the test item depends as much on the way the test item is constructed as on the content. Important topics that are tested by ambiguous, awkward, or poorly specified test items cannot be considered valid. Therefore, the selection of valid topics and the careful construction of test items are equally important parts of a valid test development process.

The conversion of a learning objective into a test item is, in part, a creative process. Yet, there are certain procedures and guidelines that can help in writing a test item and in ensuring that the test item will measure the knowledge, skill, or ability that it is intended to measure.

Five steps are involved in developing test items:

- Test item format should be determined
- Number of test items to be developed should be determined
- Skill and knowledge test items should be written
- Test item content should be validated
- Test items should be incorporated into the test bank for future use.

Determination of test item format should be guided by the action verb of the learning objective. However, any format of test item can be used to test any cognitive level if an appropriate stem is developed. Skill verbs suggest performance testing while knowledge verbs suggest one or more of the written formats. For example, "start up" and "shut down" are skill verbs that suggest a performance test, whereas "recall," "identify," and "list" are verbs more appropriately addressed by the completion and short-answer formats. Verbs such as "discriminate" and "select" should be tested using a multiple-choice format. Learning objectives that require the trainee to classify or relate are most easily tested by using the matching format. Higher-order verbs such as "synthesize" and "evaluate" may be tested using an essay format. Appendix B gives a detailed listing of the action verbs versus the domains and levels within the domains.

After the test item format is determined, the number of test items to be developed should be established. Some of the factors that significantly affect the number of test items to be developed are the number and type of examinations on which the test items will be used, the ease with which test items can be created, and how frequently a test item will be reused. The initial number of test items selected for development should range from at least one to several test items per learning objective.

Test items should be written after the format is selected and the number of test items is determined. Test items should have answer keys written at this time.

Content validation is the process by which test items are determined to be consistent between the content being learned and what is being tested. Each test item should be reviewed by subject matter experts (SMEs). These validation reviews should be documented.

A test item bank should be developed and maintained current. The test items in the test bank may be used to construct entry-level tests, progress tests, or after-instruction tests. A tracking system should be developed to correlate test items to the corresponding learning objective.

The following chapters offer guidance on how to select, construct, and score different formats of written test items (e.g., short-answer, multiple-choice, matching, and essay). Traditionally, test items that require the trainee to supply an answer (e.g., short-answer, essay) have been considered "subjective"; test items requiring the trainee to select an answer (e.g., multiple-choice, matching) have been considered "objective." If graders require subject matter expertise to interpret the answers of test takers, the test item is considered subjective. If the examination can be scored without having to interpret the answer (e.g., machine scored), it is considered objective. An objective test item is defined here as one in which: (a) there is only one correct answer, and (b) all qualified graders would agree on the amount of credit allowed for any given trainee's answer.

Writing test items that are technically sound is a skillful and challenging task. Exam writers must balance technical concepts with an array of test item writing principles to create a single "good" test item. The test item should be one that reasonably assesses what a

trainee ought to know. The test writing challenge is further complicated by the inherent ambiguities and interpretations of the written language. To make it perfectly clear in the mind of the trainee what the test item is asking for, test items must be as precise and focused as possible.

Finally, the overall written examinations can be highly streamlined if examiners adapt use of a separate answer sheet for the recording of trainee answers. The separate answer sheet affords several key benefits.

- Directions for recording answers for multiple-choice test items need only be given once (at the onset of the exam) rather than repeating "circle the answer below" for each test item
- Examiner scoring can be made easier and more rapid through either the use of an overlay key or an optical scanner
- Test item analysis can be accomplished more readily through either a single-sheet, visual profile of responses or through a computerized test item analysis via optical scanner.

In translating a learning objective into a test item, the following examples may help the examiner generate ideas for potential test items. These examples not only help in the development of test items, but they may also be useful when generating learning objectives.

1. What are the common misconceptions about _____?
2. Why is _____ important to satisfactory job performance?
3. In what sort of circumstances might it be important to understand _____?
4. What might the individual do who does not understand _____?
5. What might the consequences be of a lack of knowledge about _____?
6. How can the individual demonstrate this knowledge?

INTENTIONALLY BLANK

3. DEVELOPMENT OF WRITTEN TEST ITEMS

Guidance on the construction of each test item format is provided in the following chapters. However, there are basic principles that apply across all test item formats.

- Ensure that the concept is relevant to the ability to perform the job
- State the test item concisely
- Choose the higher cognitive level
- Make sure the test item matches the learning objective
- Omit unnecessarily difficult or irrelevant test items
- Limit the test item to only one concept or topic
- Avoid copying text directly from reference materials
- Avoid backwards logic test items
- Place the easier test items at the beginning of each section
- The test item should discriminate between those who have mastered the objective and those who have not.

Ensure that the concept being measured has a direct relationship to the ability to perform the job. The construction of the test item should clearly reflect the enabling objective. Word the test item so that it would be considered valid and reasonable to other SMEs using the same reference materials.

State the test item as concisely as possible, but provide all necessary information. The test item should be clear, grammatically correct, and free of clues to the correct answer. It should be written at a reading level appropriate for the trainee. Often the individuals who develop a test item assume that certain conditions are inherent in the question when, in fact, they are not. It is important to have others review your test items to ensure that all necessary information is included, and that all excess information is deleted. You should ask yourself: Will the trainees clearly know what they are expected to do? Do they have all the information they need to answer the test item? Does answering the test item depend on certain assumptions that must be stated?

When there is a choice between two cognitive levels, write your test item to reflect the higher level. Learning objectives and test items should be written to reflect the level of

cognitive domain that is most appropriate. Examinations should consist of higher-level cognitive test items.

Make sure that the test item matches the learning objective. It is very easy to end up with a test item that tests a relatively trivial aspect of an important learning objective. When reviewing your draft test item, ask yourself whether it is likely that someone could answer the test item correctly and still not meet the objective or perform the tasks.

Omit test items that are irrelevant. When reviewing your draft test item, ask yourself: could someone do the job safely and effectively without being able to answer the test item? If so, is it because the content is inappropriate, the wording is unclear, or the level of understanding is too great?

Limit the test item to one concept or topic, unless a synthesis of concepts is being tested. Each individual test item should be reserved for testing one topic, and that topic, as well as the intent of the test item, should be clear to both examiner and trainee. There is a common misconception that testing for multiple topics in one test item is a time-efficient way to examine. Test items containing a variety of topics only serve to confuse the trainee about the purpose of the test item and, therefore, what is expected in terms of a correct response.

Avoid copying text directly from training or other reference material. Test items written in this way generally encourage rote memorization. Further, copying from reference material can cause confusion in test items because the material lifted often draws its meaning (and importance) from its surrounding context. Therefore, important assumptions or conditions stated elsewhere in the material are often omitted from the test item.

Avoid "backwards logic" test items -- those test items that ask what should be provided in the test item, and provide what should be required in the trainee's response. It is important to test topics in a way consistent with how the topic should be remembered and used. For example, consider the following test item:

Test item:

If it takes 12.5 cubic feet of concrete to build a square loading pad 6 inches thick, what is the length of one side of the pad?

This test item gives the test takers information they should be asked to calculate, while it requires them to provide information that would be supplied in an actual work situation. In constructing your test items, make sure that you include information that trainees would typically have or have access to, and require responses that reflect the decisions, or calculations, or other information they would typically have to supply.

Place the easier test items at the beginning of each section. These test items help trainees gain composure and confidence. However, this is not to say that extremely easy test items should be included in the exam for the sole sake of relieving trainee tension.

Finally, a test item must be worded so that it discriminates between those who have mastered the objective and those who have not. A well-written test item should parallel the objective that it is testing.

INTENTIONALLY BLANK

4. SHORT-ANSWER TEST ITEMS

Short-answer items require the trainee to write a response in contrast to selecting from among a set of alternative responses. A short-answer item may be presented in such a way as to give the trainee freedom to express ideas or it may restrict a response to a given content area or answer format. This test item format is useful in testing knowledge, comprehension, and simple application. This format is more objective than essay-format test items; however, it is not free from subjectivity in scoring. The "guess-factor" is reduced from that of a multiple-choice test item because the chance of successfully guessing the answer is less. Because of varying trainee responses, the person grading the test item needs a definitive answer key and an understanding of the subject matter.

There are three common styles of short-answer test items: completion, fill-in-the-blank, and list/diagram completion. In each case, the trainee must supply the correct answer. Because the developer of a short-answer test item often has clear ideas about expected answers, it is essential that model answers, acceptable alternatives, and associated grading criteria be written at the same time as the test item.

Examples of short-answer test items are:

Completion:

How many protons and electrons does the element hydrogen contain?

Fill-in-the-blank:

In a pressurized water reactor, fuel pellets of _____ are clad in _____; the primary coolant is _____.

List/diagram completion:

Complete the blanks in the following list, indicating thermocouple type and material:

<u>TYPE</u>	<u>MATERIAL</u>
K	Alumel- _____
	Iron+ _____
	Constantan- _____
T	Constantan- _____
	Chromel+ _____
	Constantan- _____
R or S	Platinum- _____
	Rhodium+ _____

These formats are easy to construct and score. However, the disadvantages frequently outweigh the advantages, since spelling errors, grammatical errors, and legibility complicate the scoring process. Test graders cannot anticipate all possible responses and synonyms for the responses they expect, and therefore, often have to make subjective judgements on the correctness of trainees' responses.

Short-answer test items can be used to elicit various abilities. These include:

- Recalling terminology, facts, symbols, and classifications
- Applying rules and making interpretations of data and other information
- Solving scientific and mathematical problems
- Manipulating symbols and balancing mathematical and chemical equations.

5. SHORT-ANSWER TEST ITEM WRITING PROCEDURE

As with all test item formats, more consistent results are attainable if certain guidelines and rules are followed when the test items are developed. Keep the following in mind when constructing short-answer test items:

- Provide clear, explicit directions
- Avoid excessive wording or information
- Ensure the answer key matches the test item
- Avoid giving away the answer
- Avoid trick test items
- Use the question form if possible
- Word each test item in specific terms with clear meanings so the intended answer is the only one possible, and so the answer is a single word, brief phrase, or number
- Avoid copying statements word for word from texts or classroom materials
- State the precision, numerical units, or degree of accuracy expected of the answer
- Word the test item to avoid irrelevant clues or specific determiners
- Word each test item so the blank is toward the end of the sentence
- Omit important rather than trivial words from fill-the-blank test items
- Avoid "butchered" or "mutilated" sentences; use only one or two blanks in a complete sentence
- If more than one blank is used, keep the blanks equal in length.

The test item directions should provide clear, explicit directions for answering the test item so that the trainee understands what constitutes a fully correct response. Choose words carefully to ensure that the conditions and requirements of the test item are appropriately conveyed. Words such as "evaluate," "outline," and "explain," can invite lots of detail that is not necessarily relevant.

The test item developer should avoid excessive wording or information in the test item. In an attempt to make a test item operationally oriented and/or meaningful, there is a tendency

to add more information than required for a correct response. Below are two examples which illustrate this point.

First Example:

The term 'undermoderated' is used to describe one of the major characteristics of a typical reactor core. The effect that an increase in power would have on the moderator-to-fuel ratio, is to _____ the ratio and to _____ K_{eff} .

The beginning sentence of the above test item adds nothing but extra reading time.

Second Example:

Two isotopes, one with a half-life of 10 days and another with a half-life of 5 days, sit on opposite trays of a set of scales. Assuming the isotopes have identical atomic weights, and there is one Curie of each isotope, which way will the scales tilt (which weighs more)?

The situation in the above test item bears little relationship to a situation that an operator would encounter on the job.

Make sure that the answer key response matches (and is limited to) the requirements posed in the test item. The test item below only asks for a yes or no response, yet the answer key indicates that an explanation of that one-word response is required.

Test item:

Can a certified operator who has not been at the facility for a period of six months immediately resume duties in a licensed position?

Answer:

No. In any case the facility management would have to be notified before the operator resumed duties, and only after certification by an authorized representative of the facility or by a demonstration by the individual that his/her understanding of facility operations and administration is satisfactory.

The test developer should avoid giving away part or all of the answer by the way the test item is worded. In the test item below half credit is allotted for answering the first part of the test item "yes." However, a test-wise trainee can realize that the answer has to be yes, or else the second part of the test item would have read something like "If so, how? If not, why not?"

Test item:

If the letdown line became obstructed, could boration of the plant be accomplished shortly after a reactor trip to put the plant in cold shutdown?
(1 point) If so, how? (1 point)

Trick test items should be avoided. Trick test items can occur unintentionally when the answer key does not precisely match the test item. The test item below asks for how, not if, the termination criteria change.

Test item:

How does the SI termination criteria change following an SI reinitiation?

Answer:

It does not change.

Examples of short-answer test items are found in Appendix C -- Test-Item Examples.

INTENTIONALLY BLANK

6. MULTIPLE-CHOICE TEST ITEMS

Constructing a quality multiple-choice test item has two major advantages. First, the scoring of objective test items is considerably more reliable and less time consuming than scoring subjective test items. Second, since the test item requires less time to answer, more test items can be used to test trainee knowledge. This will provide better content coverage, which will also increase test validity.

There are two distinct parts of a multiple-choice test item, the *stem* and the *choices*, *alternatives*, *responses*, or *options*. The stem is the introductory statement, and its functions are to ask a question, define a task, or state a problem to be solved. The alternatives, responses, or options are the listed responses to the stem. The alternative part may be viewed in two classifications, the response that is the one correct or best answer, and the remaining incorrect responses (called *distractors*).

The multiple-choice format of test item can be a good selection to:

- To assess a trainee's ability to discriminate and make choices
- To comprehend concepts, principles, and generalizations
- To make judgements about and choices among various courses of action
- To infer and reason
- To interpret new data or information
- To apply information and knowledge.

The multiple-choice test item is not a good selection to use when the trainer needs to measure:

- The trainee's ability for recollection under minimum prompting
- The ability to explain and give examples
- The ability to produce and express unique or original ideas
- The ability to organize personal thoughts
- The ability to display thought processes or patterns of reasoning.

Multiple-choice test items require trainees to select the best answer from a series of choices. This series of choices can use calculations, descriptions, arguments, estimates, sketches, etc. They are useful for testing knowledge, comprehension, and application of knowledge and theory. Since the answer is provided, the subjective aspect of scoring is eliminated. Furthermore, scoring can be conducted by mechanical means. If the stem is properly developed, multiple-choice test items can effectively test higher-level, cognitive abilities, such as, analysis, synthesis, and evaluation. The latter abilities are difficult to measure directly through any type of written test item; however, the essay-format of test item is preferred for this application.

The multiple-choice test item is the best choice where there are numerous test items covering a category of topics, and the objectivity of scoring is the most important consideration. For instance, an examination that determines the trainee's selection for the job position. The multiple-choice test item can be used to test a greater variety of learning objectives and does not require the trainee to write out and elaborate their answers, minimizing the opportunity for less knowledgeable trainees to *bluff* or *dress up* their answers.

Multiple-choice test items should not be used when:

- The test item calls for a numerical answer, as in computational problems
- Writing the answer does not take any longer than using an answer sheet or marking the answer to the multiple-choice test item.

The development of multiple-choice test items from learning objectives involves three steps:

- Create the stem of the test item by forming a question of an incomplete sentence that implies a question
- Write the correct answer to the stem in as few words as possible
- Write distractors that are plausible to trainees lacking the degree of knowledge you want the test item to assess.

The following table provides tips for use in developing the stem of a multiple-choice test item.

TABLE 1. Tips for developing the stem of a multiple-choice test item

To do	To avoid
<ol style="list-style-type: none"> 1. If possible, write as a direct question. 2. If an incomplete sentence is used, be sure that <ul style="list-style-type: none"> • It implies a direct question • The alternatives come at the end (rather than in the middle) of a sentence. 3. Control the wording so vocabulary and sentence structure are at a relatively low and non-technical level. 4. In test items testing definitions, place the word or term in the stem and use definitions or descriptions as alternatives. 	<ol style="list-style-type: none"> 1. Avoid extraneous, superfluous, and non-functioning words and phrases that are mere <i>window dressing</i>. 2. Avoid negatively worded test items. 3. Avoid phrasing the test item so that the personal opinion of the trainee is an option. 4. Avoid textbook wording. 5. Avoid <i>cluing</i> or <i>linking</i> test items (i.e., having the correct answer to one test item clued or linked to the correctness of the answer of a previous test item).

6.1 Styles of Multiple-Choice Test Items

Multiple-choice test items are the most common form of structured response-type test items. By definition, multiple-choice includes alternate-choice (true/false, yes/no, right/wrong) and matching. Both types vary only in the number of alternatives available.

Matching test items can be used where developers want to assess ability to draw relationships between a similar set of ideas. While matching test items can be objectively scored, care must be taken to ensure that associations made in each column are similar in nature and kind (i.e., conditions and procedures, causes and results, actions and individuals).

The most common multiple-choice form consists of a stem and the choices. The choices include the correct answer and several distractors. This form is illustrated in the following example:

Test item:

The atoms of a chemical element all have the same	STEM
(a) Atomic number (Correct choice)	C
	H
(b) Relative atomic mass (Distractor)	O
	I
(c) Number of neutrons in the nucleus (Distractor)	C
	E
(d) Number of nucleons in the nucleus (Distractor)	S

There are various ways of producing and selecting distractors. The test item developer can independently (or in discussion with colleagues) determine possible distractors. An effective method is to develop a simple test item of a fill-in-the-blank statement. For example: What do all the atoms of a chemical element have in common? Then, give this test item to a group of trainees. Possible distractors that will emerge from their responses are usually more effective than those devised by the test-item developer, particularly since the distractors also identify trainee learning errors.

The selection of distractors should be based on the required depth of trainee understanding. One set of distractors may be appropriate only when a general understanding of a subject is required, while a different set of distractors is needed when an in-depth understanding is required (see the Model A example below).

Developing a good multiple-choice test item, written at the application level of the cognitive domain, which also adheres to good test item construction guidelines (i.e., valid, operationally oriented, plausible distractors) presents a challenge to the developer. To assist and guide you in designing multiple choice test items, the following four models are presented:

Model A:

- (a) Correct answer
- (b) Incorrect answer
- (c) Incorrect answer
- (d) Incorrect answer.

Model A depicts the traditional multiple-choice design format. This model shows one correct single word/phrase answer followed by three incorrect single word/phrase options. The length of all options should be similar.

Model B:

- (a) Correct answer
- (b) Plausible misconception
- (c) Incorrect answer
- (d) Incorrect answer.

Model B is a variation of Model A where a plausible misconception is used as an incorrect answer. Again the length of all options should be similar.

Model C:

- (a) Correct answer with correct condition (e.g., because, since, when, if, etc.)
- (b) Correct answer with incorrect condition
- (c) Incorrect answer with incorrect condition
- (d) Incorrect answer with incorrect condition.

Model C depicts an acceptable multiple-choice design that uses conditions with answers (a condition in a setting, event, cause/effect that may make the answer correct or incorrect). Notice that Model C shows only one correct answer with its correct condition.

Model D:

- (a) Correct answer
- (b) Incorrect answer
- (c) Correct answer with incorrect condition
- (d) Incorrect answer with incorrect condition.

Model D illustrates an acceptable model when it may not be possible to create all options in uniform length. This model shows paired lengths -- two long and two short options -- which avoids setting any single option apart (either too long or too short) from the remaining options.

The four models presented are basic. Other models may be developed and used in combination with one another.

6.2 Examples of Multiple-Choice Test Items

Four-Option Multiple-Choice Test Item -- A multiple-choice test item with the same correct answer (key) can be written in several ways depending on the breadth of knowledge being assessed. The following example shows three variations of a multiple-choice test item. (The correct answer in each is "rad," but the distractors have been chosen with different objectives in mind):

Variation 1:

In health physics, the unit of absorbed dose is:

- (a) Volt (b) Rad (c) Calorie (d) Newton

The above set of choices covers such a wide range that discrimination would be extremely easy for those with only a vague understanding of health physics.

Variation 2:

In health physics, the unit of absorbed dose is:

- (a) Rad (b) Rem (c) Kerma (d) Disintegration/second

The above set of choices covers such a narrow band of knowledge that only those who have a good understanding of health physics would choose the correct answer.

Variation 3:

In health physics, the unit of absorbed dose is:

- (a) Counts/second (b) Rad (c) Half-life (d) Rem

The above set of choices covers two bands of knowledge. Counts/second and half-life would be immediately dismissed by those with some understanding of health physics, leaving the choice between (b) and (d) (i.e., one choice from two, not four, distractors).

7. MULTIPLE-CHOICE TEST ITEM WRITING PROCEDURE

Although multiple-choice test items are not as easy to construct as other forms, they are very versatile, can be used to test all levels and types of knowledge, and minimize the likelihood of the trainee obtaining the correct answer by guessing. The following guidelines should be considered as you construct multiple-choice test items:

- Use four-answer options
- Do not use "none of the above" or "all of the above"
- Do not present a collection of true-false statements as a multiple-choice test item
- Define the test item, task, or problem in the stem
- Avoid negative stems
- Provide sufficient counterbalance in the distractors
- Include common misconceptions as distractors
- Make answers plausible and avoid overlapping answers
- Ensure that the distractors grammatically follow the stem.

The four-distractor, multiple-choice test item is the most common and is the style preferred. The five-answer option contributes nothing to the test item but confusion, and any format with fewer than four distractors makes guessing correctly more probable.

"All of the above" test items provide inadvertent clues to the trainee. When the "all of the above" option is the correct response, the trainee need only recognize that two of the options are correct to answer the test item correctly. When the "all of the above" option is used as a distractor, the trainee needs only to be able to determine that one option is incorrect in order to eliminate this option. "None of the above" responses should not be used with "best answer" multiple-choice test items, since it may always be defensible as a response.

Each test item should be focused on one topic. A test item containing answer options related to many separate issues does not increase the efficiency of the test item. To the contrary, test items with multiple topics only confuse the trainee about the meaning and purpose of the test item.

The test item, task, or problem should be defined in the stem of the test item. Include as much information about the problem or situation in the stem as necessary, leaving only the solution, action, or effect for the answer options:

Poor:

At 50% power

- (a) The equilibrium xenon reactivity worth is approximately equal to the equilibrium xenon worth at 100% power.
- (b) The equilibrium xenon reactivity worth is approximately one-half the equilibrium xenon worth as 100% power.
- (c) The equilibrium xenon reactivity worth is approximately two-thirds the equilibrium xenon worth at 100% power.
- (d) The equilibrium xenon reactivity worth is approximately three-fourths the equilibrium xenon worth at 100% power.

Better:

Which of the following approximates the 50% power equilibrium xenon reactivity worth to the 100% power equilibrium xenon reactivity worth?

- (a) Equal to
- (b) One-half
- (c) Two-thirds
- (d) Three-fourths.

Negative stems should be avoided because they are unnecessarily difficult to interpret. Negatively stated stems can also lead to double negatives. It is more important to test what something is than what it is not. If a negative stem is necessary, highlight the negative word (e.g., not, never, least).

It is very tempting to write negatively stated test items, since they can be constructed by picking three true statements out of the reference material and changing a fourth statement to make it false. However, studies have shown that trainees do not do as well on negatively stated test items, either because they overlook the negative word or because negatively stated test items require trainees to pick an answer or characteristic that is not true, which can be somewhat confusing. In addition, these test items tend to emphasize negative learning. For example, consider the following stem of a multiple-choice test item:

Test item:

During 100% power operation, the feedwater dump valve opens inadvertently.
The condensate pumps will not do which of the following?

This stem can be made to read positively as follows:

Test item:

During 100% power operation, the feedwater dump valve opens inadvertently.
The condensate pumps will:

- (a) Increase flow to maintain feedwater flow rate.
- (b) Trip due to a runout condition.
- (c) Have no response.
- (d) Trip due to low suction pressure.

There are times when a negatively stated test item is unavoidable. However, never use a negatively stated stem with a negatively stated answer option:

Test item:

Which of the following indications would not be expected and might indicate an instrument failure?

- (a) The CRD 'travel' lamp does not indicate when Group 8 rods are in motion.
- (b) Group 7 out-motion is prevented past 91.4%.
- (c) When you depress the 'CRD travel in' lamp-test push button, the 'CRD travel out' lamp comes on.
- (d) During a transfer of a group from direct current, hold to auxiliary; when you select 'SEQ-OR,' the 'SEQ-OR' lamp is on and the 'SEQ' lamp goes off.

Notice how confusing the option (a) answer is in combination with the test item stem.

Provide sufficient counterbalance in test items with multipart answers. Multiple-choice test items can legitimately contain multipart answer options. However, if the answers contain too many parts and/or too many options for each part, cues indicating the correct answer may be unavoidable. Consider the following example:

Test item:

The RCS is in hot standby with no reactor coolant pumps running. If OTSG pressure is decreased, according to the plant verification procedure, which of the following temperature responses indicate the presence of natural circulation?

- (a) T_h increases, T_c remains the same.

- (b) T_h increases, T_c decreases.
- (c) T_h decreases, T_c decreases.
- (d) T_h remains the same, T_c decreases.

The trainee could choose the correct answer (c) without knowing about the T_c temperature response in this situation, since " T_h decreases" only occurs in option (c).

Notice that two-part answers, with each part containing a two-option response, provides complete counterbalance, since all contingencies can be covered in four responses. For example:

Test item:

Which of the following is a definition of quadrant power tilt ratio?

- (a) Minimum upper detector output divided by average upper detector output.
- (b) Maximum upper detector output divided by average upper detector output.
- (c) Minimum upper detector output divided by average lower detector output.
- (d) Maximum upper detector output divided by average lower detector output.

A multipart test item which is highly recommended is one in which the two-part answer options consist of a two-level response (e.g., yes/no; off/on) and a reason. For example:

Test item:

Which of the following best describes the behavior of equilibrium xenon reactivity over core life?

- (a) It decreases, because of the increased fuel burn up.
- (b) It decreases, because of the decrease in plutonium-xenon yield.
- (c) It increases, because of the increase in thermal flux.
- (d) It increases, because of the decrease in boron concentration.

When possible, include common misconceptions as distractors. Since the purpose of the examination is to differentiate between competent and less-than-competent trainees, a good source of test items involves topics in which there are common misconceptions about important knowledge topics. For example, the following test item was based upon a common misconception about loss of subcooling margin:

Test item:

During a LOCA with a resultant loss of subcooling margin, why are the reactor coolant pumps (RCPs) secured?

- (a) To prevent pump damage resulting from operation under two-phase conditions.
- (b) To prevent core damage resulting from separation upon subsequent loss of RCS flow.
- (c) To reduce RCS pressure by removing the pressure heat developed by the RCPs.
- (d) To remove the heat being added to the RCS by the operating RCPs.

Make all answer options homogeneous and plausible. For example, the choices of answers to the following test item illustrates these characteristics:

Test item:

On a loss of condenser circulating water intake, the upper surge tank, hotwell, and condensate storage tank will supply sufficient feedwater to allow decay heat removal for approximately:

<u>Poor</u>	<u>Better</u>
15 minutes	8 hours
8 hours	24 hours
48 hours	48 hours
3 months	72 hours

One method of changing the difficulty level of a test item is to vary the similarity among answer options. Develop distractors that are similar enough to be chosen by those who do not meet the testing objective, yet different enough so that they do not test trivial issues or distinctions. If the answer options have a logical sequence, put them in order.

Avoid overlapping answer options:

Test item:

The SPND uses rhodium which decays with a half-life of 42 seconds. How long will it take for a detector to indicate approximately 95% of an instantaneous power level change?

<u>Poor</u>	<u>Better</u>
2-4 minutes	1-2 minutes
4-6 minutes	3-4 minutes
6-8 minutes	5-6 minutes
8-10 minutes	7-8 minutes

To prevent giving the correct answer, ensure that the distractors do follow grammatically from the stem -- for example:

Test item:

During 100% normal power operation a single steam flow element to the steam generator feedwater fails high. The steam generator feedwater control system will cause:

- (a) The feedwater valves to increase steam generation level slightly before returning the level to normal.
- (b) Before returning the level to slightly above normal, the feedwater valves to increase the steam generator significantly.
- (c) The feedwater valves to increase the steam generator level to the level of a reactor trip.
- (d) The feedwater valves to increase the steam generator level slightly and maintain the increased level.

Test item with Improved Distractors:

During 100% normal power operation a single steam flow element to the steam generator feedwater fails high. The steam generator feedwater control system will cause:

- (a) The feedwater valves to increase steam generator level slightly before returning the level to normal.
- (b) The feedwater valves to increase the steam generator level significantly before returning the level to slightly above normal.

- (c) The feedwater valves to increase the steam generator level to the level of a reactor trip.
- (d) The feedwater valves to increase the steam generator level slightly and maintain the increased level.

INTENTIONALLY BLANK

8. MATCHING TEST ITEM WRITING PROCEDURE

With matching test items, a trainee is required to match each word, sentence, or phrase in one column with a word, sentence, or phrase in another column. The test items in the first column are called "premises." The answers in the second column are called "responses." Consider the following guidelines when constructing matching test items.

- Give specific directions for each matching test item that indicate on exactly what basis the matching is to be done.
- State in the directions whether a response can be used only once or more than once.
- Only use logically related material in a single matching test item -- nothing should be listed in cross-related columns that is not part of the subject of the test item.
- Consider the following possibilities for pairing premises and responses:

<u>PREMISES</u>	<u>RESPONSES</u>
Terms or words	Definitions
Short test items	Answers
Symbols	Proper names
Causes	Effects
Principles	Situations in which they apply

- For any single matching test item, use only one set of premise-response formats (e.g., only terms with definitions).
- Each response should be a plausible answer for each premise -- for example:

Test item:

Match the alarm (letter) with its appropriate tone (number). Responses (numbers) may be used more than once:

<u>PREMISES</u>	<u>RESPONSES</u>
Fire	Pulse tone ___
Reactor building evacuation	Wailing siren ___
Site evacuation	Steady tone ___

- Present an unequal number of premises and responses, or allow responses to be used more than once -- this will inhibit the trainee from obtaining clues to a correct answer through a process of elimination.
- Arrange the responses in a logical order, such as alphabetical or numerical.
- The entire matching test item should contain no more than six premises.
- Place the entire matching test item on one page so the trainee need not flip back and forth between pages.

Directions to matching test items should not require more than one response to a premise. This practice not only fragments scoring and alters the implicit equality of the premises, but also may encourage "shotgunning," the writing in of all responses to a premise in order to receive some partial credit.

If partial credit is possible, establish how it will be assigned in the examination key. For example, a five-premise matching test item might have 20% credit for each correct match.

9. ESSAY-FORMAT TEST ITEMS

Test items that are called essay-format include those requiring the trainee to provide calculations, graphs, descriptions, arguments, estimates, sketches, or combinations of these. The trainee has to determine the answer and select the means of communication. Essay-format test items should be used when the need is to have the trainee express himself/herself or to recall information from memory rather than to present the information to the trainee for recognition or identification. Test items of this format can test the whole range of abilities, but they usually are considered most suitable for application, analysis, synthesis, and evaluation. The test item and the associated grading criteria must be prepared concurrently. The examiner grading the test item must be knowledgeable in the subject matter.

9.1 Styles of Essay-Format Test Items

There are two major styles of essay-format test items: structured and unstructured. In each case, trainees must supply the correct answer. However, the unstructured test items require a greater amount of analysis and composition than are required by structured test items.

The unstructured essay-format test item is appropriate where a learning objective expects a trainee to organize ideas, to develop a logical argument, to present evaluations of certain thoughts, to communicate thoughts and feelings, or to demonstrate other abilities requiring original written expression. An example of this format would be an assignment for the trainee to "write a paper explaining the pro's and con's of using written procedures." A disadvantage of the unstructured essay-format test item is that scoring the response in an objective manner is difficult and time-consuming.

The structured essay-format test item does not have to be limited to testing for recall and comprehension. It can be designed to require the trainees to apply their knowledge and skills to solve new problems or analyze a novel situation. One way to do this is to write test item statements that require the trainee to apply their knowledge to specific material given in the test. For instance, a description of a particular, along with the initial facility conditions, can be given in the test item statement. One advantage of the structured

essay-format test item is that by narrowing the focus of the test item statement to elicit specific and well-defined behaviors, the trainees are more likely to interpret the statement as the test developer intended. The structured essay-format test item also prompts the test developer to provide more clarity in presenting the correct answers, which can lead to improved reliability of scoring.

A disadvantage of all essay-format test items lies in their application, since they are often easier to develop than the multiple-choice test item. A good application of the essay-format test item is to offer the trainee an opportunity to exhibit abilities for written production, organization, expression, and interrelationships among ideas. A poor use of the essay-format test item is to test for factual information. The trainer can make a more appropriate test item format selection by analyzing whether the related objectives require the trainee to *recall and supply* answers or to *recognize* already supplied answers. If the latter applies, the multiple-choice test items are preferable because they allow for more objective scoring.

9.2 Examples of Essay-Format Test Items

It is possible to express the same test item as either a structured or unstructured test item. Two extremes are shown here:

Structured Essay-Format Test items:

A pressurized water reactor at full power has the following conditions:

- (a) Core flow (100%)
- (b) T_{avg} (588°F)
- (c) Core ΔT (60°F)
- (d) Steam Generator ΔT ($T_{avg} - T_{steam}$) (44°F).

The reactor is operating at steady-state power with four loops operating. One reactor coolant pump trips, resulting in a three-loop operation. The rod control system is in automatic. The reactor has not tripped.

- (a) Sketch graphs of RCS mass flow, reactor power, core ΔT , and steam generator pressure versus time if no operator action is taken. Use the same time axis for all graphs. (4 points)

- (b) Calculate the final reactor power, T_{avg} , core ΔT , and steam generator ΔT . (4 points)

This test item gives the trainee clear guidelines on what is expected. This is termed a structured test item and requires little personal initiative, but a clear requirement exists to apply knowledge gained from theoretical and practical experience in the course. Also, this test item lends itself to more consistent grading, and is, therefore, more objective.

Unstructured Essay-Format Test item:

A pressurized water reactor is operating at 100% steady-state power. One reactor coolant pump trips resulting in a three-loop operation. Explain what happens if the reactor does not trip and no operator action is taken. State any assumptions you make. (8 points)

This test item requires the trainee to interpret the word "explain." S/He must decide what the important issues are, what assumptions to make (possibly based on parameters taken from his/her own experience) and then communicate the issues and assumptions by whatever means he thinks are appropriate. Some trainees may use graphs and a short explanation; others may use only an explanation. Each answer will be unique. This test item requires the trainee to clearly express understanding of the topic.

Unstructured essay-formats are more difficult to grade consistently because of the variations that are possible, relative to assumptions the trainee must make. Consequently, the grading becomes much more subjective in test items such as the one in the foregoing example.

For assessment purposes, most test-item developers prefer to write structured essay-format test items. Unstructured essay-format test items are most appropriate in learning exercises

in which feedback and additional clarification can be provided if necessary (oral examination, on-the-job training, walk-through, etc.).

10. ESSAY-FORMAT TEST ITEM WRITING PROCEDURE

The layout of the test item and the clarity of the wording are crucial for the trainee to understand what is required when essay-format questions are used. Generally the words "how" and "why" are the best, followed by "what" and "when," although the latter two can more easily lead to a single word or very short answers.

Trainees find particular difficulty with interpretation of words such as "discuss," "compare," and "explain." The level of the answer expected is important in determining whether a more specific statement is useful.

The test item developer should follow a systematic method such as the following:

- Define the behavior the trainee is expected to exhibit or describe the process to be exhibited before beginning to write the essay-format test item.
- Allow adequate time for developing the test item, model answer, and grading criteria. Essay-format test items usually require several revisions during the development process. Consider stating the anticipated answering time for the test item.
- Referring to the learning objectives, select a topic from which the test item is structured. Ask only for important relevant information.
- Ensure the test item is within the educational maturity level of the trainees.
- Determine the level of learning expected in response to a test item on this topic from the learning objective. Use the words "discuss," "explain," or "compare" only when deliberately aiming at higher cognitive levels.
- Define the problem explicitly -- an essay-format test item is of no use if the trainee cannot comprehend the test item. It must be open to only one interpretation so all trainees address the same topic.
- Ask test items that require the trainee to demonstrate the ability to use essential knowledge and do so in situations that are new or novel for the trainee, rather than simply recalling information.
- Limit the problem -- the scope of a content area should not be too large. Remember that trainees will take longer to answer the test item than the developer.

- Word the test item so the trainees know the limits of the tasks, their purposes, and can answer them in the time allotted.
- Ask questions that are relatively specific or focused, and which require relatively brief responses.
- If additional guidance for a particular test item is necessary beyond that shown in the general instruction, make a clear and concise statement.
- The subject matter experts should also agree on the relevance or importance of a particular test item to course completion, as well as the correct response.
- Develop the test item grading criteria. Identify the key responses expected from the trainee. Identify each necessary response (e.g., by underlining key words, phrases, or steps, or by circling important parts of graphs or diagrams). Determine the percentage value to be allocated for each test item subpart (based on their relative importance).
- Check the test item against the grading criteria:
 - Does the test item clearly request what is considered important in the grading criteria?
 - Is the time necessary to answer the test item proportional to the importance of the test time? (Critical objectives justify more time if required.)
 - Do trivial details (e.g., an extended calculation) take too much time to address?
 - How could the test item and/or the grading criteria be revised to be more consistent?
- Ensure that the sequence of requirements is in a logical order.
- Ensure that all the data (stimulus material) needed by the trainee is provided and clearly displayed together.
- Avoid test items that can be answered by a single word or short phrase.
- Avoid negative statements in test items; they may lead to double negatives that are difficult to interpret.
- State clearly whether exact answers or estimates (precise figures or sketches) are required.
- Ensure that the point value for each test item or part of a test item is stated.
- Write the test item on a form such as that shown in Appendix A -- Test Item Development Form.

DOE-HDBK-1204-97

- Word the test item so that trainee can judge the approximate length of the answer desired and knows the point values or weight each will be given.
- Have another test item developer review the test item for accuracy, format, and readability.
- Word the test item so the SMEs can agree on the correctness of a trainee's response.

INTENTIONALLY BLANK

11. TEST ITEM ANSWERS AND POINT VALUE

For each test item, a correct answer should be supplied in order to develop an answer key. The answers should be unambiguous so that the grading of the test item will be consistent. The minimum response for full credit must be given, and indications of the relative value of partial responses should be made. This kind of information is obvious in multiple-choice and matching test items, but short-answer and essay test items do leave room for interpretation.

To prevent interpretive inconsistencies in grading short-answer or essay test items, the designated instructor should include the minimum answer required for full credit. If the exact wording or specific concept must be included to fulfill the requirement of a correct answer, one method that may be used is to underline the necessary parts and assign point value to the individual parts. If the exact underlined wording is not necessary, include a note to the grader that the content is necessary, but the wording is left to the grader discretion. For example:

Test item:

State the purpose of the emergency diesel generator. (4 points)

Answer:

Provides emergency power (2.0 points) to essential loads (1.0 point) in the event of a total loss of power (1.0 point). NOTE: *Alternate wording acceptable—grader discretion requested.*

After the test item and answer key are written, the point value should be assigned in a clear and consistent manner. The assignment should be based on a point scale with the "easiest" test item assigned lower point value and the "more difficult" test item assigned higher point value. The point value should be placed in parentheses at the end of the test item stem.

11.1 Cognitive Level vs. Point Values

To determine ease or difficulty of a test item, consider the criticality of the enabling objective that is being tested to the overall mastery of the terminal objective which will allow successful performance of the job task. Also consider the cognitive level at which the

objective being tested is written. For example, if you are writing multiple-choice test items for three enabling objectives, two of which are written at the knowledge level of Bloom's Taxonomy to support a third which is written at the comprehension level, then the point value assigned to the higher cognitive level would be higher than the point value assigned to the lower cognitive level. For example, this is how these concepts might be applied:

<u>ACTION VERB</u>	<u>COGNITIVE LEVEL</u>	<u>POINTS</u>
E.O. 1.1 List	Knowledge	(2.0 pt)
E.O. 1.2 State	Knowledge	(2.0 pt)
E.O. 1.3 Determine	Comprehension	(4.0 pt)

If multiple-choice test items are written for these action verbs, the first two test items would have a lower point values than the third. This is because the third is more "important," more critical to the mastery of the terminal objective (the first two objectives support the third). Also, the third action verb is at a higher cognitive level (comprehension) than the cognitive level of the first two (knowledge).

Other considerations for determining point value of test items include the following:

- Impact on successful performance of the job task if the answer is not known by the trainee
- Learning level required (lowest level being knowledge and the highest level being evaluation)
- Number of answers required for the test item
- Relative degree of difficulty.

Another consideration in the assignment of point value is related to the test item. Test items having multiple parts requiring answers should have at least one point per answer, and matching test items should have at least one point per match. Point value allocations should be made and approved when drafted test items are submitted for review and approval.

11.2 Scoring Essay-Format Test Items

A major difficulty in the use of essay-format test items is the inability to maintain consistent scoring. The answer must clearly identify the key points of each test item, and the grading criteria must be followed closely to ensure consistency of scoring. Effective grading criteria eliminate bias toward trainees (whose answers appear plausible, have neat handwriting, or are known to the instructor) by removing subjectivity and promoting objectivity.

During scoring, it is easy to drift from the established grading criteria. It is, therefore, necessary to check each answer against the answer and grading criteria for every test. It is also advisable for the same person to score one test item across all tests before scoring the next test item. This increases the consistency of the scoring process.

Also during scoring, the examiner may find that a large percentage of trainees incorrectly answer a test item. The examiner should then reevaluate the test item to determine if the test item could have been interpreted differently than what was intended or if there may have been other potentially correct answers.

INTENTIONALLY BLANK

12. USE OF HANDOUTS

The way in which handouts are incorporated into the exam has a bearing on the reliability of the examination. If a trainee has to search for the applicable drawing or curve, it adds to their frustration level and the time consumed in test administration. The format of the exam itself also contributes to the exam's reliability. Not only does it take more time to complete an examination that incorporates a multitude of administrative requirements, but also, it detracts from the trainee's ability to focus on just the exam test items.

Include handouts such as matrices or incomplete system drawings on which the trainee must transcribe his/her answer directly behind the page of the applicable test item. This will help ensure the integrity of the trainee's package of answers.

Include all other handouts, such as reference material, Technical Safety Requirement excerpts, etc., within a separate handout package. This package should be page checked with the trainees before exam administration. Booklets, rather than copied pages, should be used whenever possible.

Use of single-page/single-test item layout should be considered. This has proven to facilitate both the exam grading and the test taking processes. It also minimizes the administrative burden on the trainee, greatly reducing the chances of any test items being missed.

INTENTIONALLY BLANK

13. OPEN-REFERENCE TEST ITEMS

Training environments use several teaching aids that promote trainee learning. Many of these aids (e.g., textbooks, manuals, and student guides) assist the trainee in mastering the required subject matter during the initial training phase. The trainee is expected to comprehend the material and commit important facts to memory. However, there are many references (e.g., tables, charts, schematics, and procedures) trainees do not need to commit to memory (and should not) but have to interpret while performing their job.

To test trainee competence in the use of this material, test items that require the use of references to solve a problem or to reach a conclusion should be developed. This requires that the reference, or a sufficient subset of the reference, be provided to the trainee during administration of the test. This use of reference materials during testing is referred to as open-reference or open book testing.

The test developer should determine which references and their applications are necessary after reviewing the learning objectives and the test specifications. While the open reference test is essentially no different than other written tests, there are several points to consider when using this method.

- References are considered tools trainees use to solve a problem or reach a conclusion. The test items should not directly test knowledge of the references (this should be done in a closed book section); rather, it should test for proper use and application of the references.
- Open-reference test items should not be a "direct lookup" in a reference. For example, asking the specific setpoint for a protective action or the horsepower of a motor are direct lookups. Rather the reference material should be an aid in going through a series of knowledge-based decisions in order to arrive at a logical conclusion or answer.
- Open-reference test items should test the trainee's ability to locate, use, and apply the information found in the references.

- Test items should contain unique or varied circumstances that the trainees have not previously encountered. This makes the test item a true indicator of the trainee's ability to apply knowledge through the use of the references versus merely remembering an application from an earlier training session. Familiarity with routine applications of a reference can lower the learning level to simple recall.
- If a test item has references provided, it is not appropriate to make the test item more or less difficult than a comparable closed book test item. While very difficult test items may be useful in differentiating among the most able trainees, they are not appropriate for job or task qualification. Do not make an open reference test item easier or more difficult than other formats.
- Keep test item requirements as close to real life situations as possible. It is important the test items address the use of references in a context similar to that found in the job environment. While providing a chart from a handbook is good, giving the trainees handbooks and requiring that the chart be found can be even better. Keep the content of open reference test items valid.
- Care must be taken in the development of open-reference test items. Under certain circumstances, these may become double jeopardy test items. In other words, if the test developer is testing the ability to use a graph vice locate it, the trainees will not be able to demonstrate use of the graph if they cannot find it.

An open-reference test item is the most challenging kind of test item to develop because it should challenge the highest cognitive level of the trainee. These test items are the result of a great deal of effort on the part of the developer. Several revisions and reviews by others are usually required to refine the test item into a quality test item that will challenge the trainee's decision making ability.

APPENDIX A
TEST ITEM DEVELOPMENT FORM

INTENTIONALLY BLANK

TEST ITEM DEVELOPMENT FORM

(Test Item Reference Number)

SIGNATURE DATE

Developed by:

Test Item Format:

Short Answer _____

Reviewed by (SME):

Multiple Choice _____

Essay _____

Other _____

Reviewed by (IT):

Approved by:

Purpose of test item (i.e., objectives tested, areas covered, content tested): _____

TEST ITEM DEVELOPMENT FORM

Reference to course material (which lecture/lesson plan, what references, page numbers):

Test item statement (provide complete test item with point values): _____

Test item response (model answer and grading criteria, including partial credit): _____

Alternate acceptable answers: _____

APPENDIX B
ACTION VERBS

INTENTIONALLY BLANK

APPENDIX B

ACTION VERBS

TABLE 1. Action verbs versus level for the cognitive domain.

Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
define	convert	apply	analyze	arrange	appraise
describe	defend	associate	appraise	assemble	assess
identify	describe	build changes	break down	categorize	choose
label	diagram	calculate	classify	collect	compare
list	discuss	choose	compute	combine	conclude
locate	distinguish	appropriate	conclude	compile	contrast
match	estimate	procedures	contrast	compose	criticize
memorize	expand	collect	criticize	construct	critique
name	explain	information	debate	create	decide
outline	express	compute	determine	design	describe
recall	extend	construct	diagram	develop	develop
recite	generalize	demonstrate	differentiate	devise	criteria
recognize	given examples, identify	discover	discern	explain	discriminate
record	illustrate	dramatize	discover	forecast	estimate
relate	infer	employ	discriminate	form	evaluate
repeat	interpret	find solutions	distinguish	formulate	explain
reproduce	locate	illustrate	examine	generalize	interpret
state	measure	interpret	experiment	generate	judge
select	outline	locate	generalize	hypothesize	justify
tell	paraphrase	manipulate	identify	manage	make
	predict	modify	illustrate	modify	judgement
	recognize	operate	infer	organize	measure
	reconstruct	perform	inspect	plan	rate
	report	practice	inventory	predict	revise
	restate	put in operation	locate	prepare	score
	review	relate	note	produce	select
	rewrite	repair	organize	propose	summarize
	summarize	schedule	outline	rearrange	support
	translate	shop	point out	reconstruct	
		show	question	relate	
		sketch	relate	reorganize	
		solve	select	revise	
		translate	separate	rewrite	
		write	solve	set up	
			subdivide	structure	
			test	summarize	
				tell	
				theorize	
				write	

ACTION VERBS**TABLE 2.** Action verbs versus level for the affective domain.

Awareness	Reinforcement	Promotion	Defense
ask accept accumulate attends to be aware of choose combine control describe differentiate follow reads receive recognize reply responds select separate set apart share	acclaim adhere applaud approve augment commend comply conform discuss follow obeys play practice praise volunteer	accepts adhere advocate alter argue assist encourage follow help initiate justify model prefers propose subsidize select support	abstract act argue arrest avoid balance debate defend define discriminate display formulate influence intervene manage organize prevent resist resolve

APPENDIX B

ACTION VERBS**TABLE 3.** Action verbs for the psychomotor domain.

The Psychomotor Domain				
acknowledge	compare	heat	override	shutdown
activate	complete	heat up	perform	sketch
actuate	compute	hoist	plot	splice
add	connect	hold	position	spray
adjust	control	identify	prepare	start
align	cool	immerse	pressurize	start up
alternate	correct	increase	prime	steer
analyze	construct	inform	print	stop
apply	cut	inspect	pull	store
assemble	couple	install	pump	switch
assess	decrease	insert	push	supply
assist	deenergize	isolate	purge	synchronize
backwash	depress	investigate	rack in/out	tagout
balance	deselect	jog	raise	test
begin	detect	letdown	reactivate	throttle
bleed	dilute	lineup	read	titrate
block	direct	load	rebuild	trace
boil	disassemble	locate	recirculate	track
borate	disconnect	lock	record	transfer
build	display	lower	regulate	transmit
bypass	dispose	lubricate	release	transport
calculate	dissolve	maintain	remove	trip
calibrate	don	maneuver	repair	tune
call	draft	manipulate	replace	turn
center	draw	measure	return	type
change	energize	mix	rinse	unlatch
charge	enter	monitor	run	unload
check	establish	move	sample	unlock
choose	estimate	neutralize	scan	uncouple
circulate	exit	observe	secure	vent
clean	feed	obtain	select	warm
clear	fix	open	sequence	warmup
close	flush	operate	service	weigh
code	formulate	organize	sharpen	weld
collect	guide	overhaul	shut	withdraw
				zero

INTENTIONALLY BLANK

APPENDIX C
TEST ITEM EXAMPLES

INTENTIONALLY BLANK

TEST ITEM EXAMPLES

C.1 Short-Answer Test Item Examples

Satisfactory:

Test item:

Define the following terms:

- (a) Minimum critical power ratio (MCPR)
- (b) Rod density
- (c) Maximum fraction limiting power density (MFLPD).

Answer:

- (a) MCPR is the smallest critical power ratio that exists in the core.
- (b) Rod density is the number of control rod notches inserted as a fraction of the total number of notches.
- (c) MFLPD is the largest value of fraction limiting power density that exists in the core.

Unsatisfactory:

Test item:

Select the words from the list provided to complete the simplified descriptions. Place the letter corresponding to each answer in the blanks. Each correct response is worth two (2) points.

- (a) The _____ critical power ratio (MCPR) is the smallest CPR that exists in the core.

TEST ITEM EXAMPLES

- (b) Rod density is the number of control rod notches _____ as a fraction of the total number of control rod notches. One hundred percent rod density is achieved when all rods are _____.
- (c) The _____ fraction of limiting power density is the _____ value of FLPD that exists in the core.

Word List:

- (a) Maximum (c) Smallest (e) Inserted
(b) Minimum (d) Largest (f) Withdrawn

Answer:

- (a) 2
(b) 5,5
(c) 1,4.

Comment:

- Since the alternatives are provided, this is actually a matching test item.
- Completion test items should not be used if the intent is to test complex knowledge.
- The answer choices are identified by numbers not letters (does not match instructions).
- The word list includes synonymous distractors (e.g., maximum versus largest, smallest versus minimum).
- The acronym MCPR identifies the correct answer as beginning with "M."

TEST ITEM EXAMPLES

Satisfactory:

Test item:

Procedure AOP-2-1909, 'Inability to Shut Down with Control Rods,' states that when certain conditions are present, the shift supervisor is required to use the standby liquid control system to shut down the reactor.

List the four conditions identified in Procedure AOP-2-1909 that must be considered by the shift supervisor in determining the need to shut down the reactor using the standby liquid control system. (4 points)

Answer:

- (a) Five (5) or more adjacent control rods are not inserted past the 06 position.
- (b) Thirty (30) or more total control rods are not inserted past the 06 position.
- (c) Reactor pressure vessel level cannot be maintained.
- (d) Suppression pool water temperature cannot be maintained below 110°F.

Unsatisfactory:

Test item:

Procedure AOP-2-1909, 'Inability to Shut down with Control Rods,' states that '...if at any time, either condition b(1) or b(2) exists, and cannot be shut down and, either (a.1) or (a.2), and if it is obvious that within the reactor, in the judgment of the shift supervisor (or in his absence, a licensed operator), a hazard exists to the environs, personnel or the plant, use the standby liquid control system per AOP-2-1400....' (4 points)

- (a) List conditions (a.1) and (a.2).

TEST ITEM EXAMPLES

(b) List conditions b(1) and b(2).

Answer:

(a.1) RPV level cannot be maintained. (1.5 points)

(a.2) Suppression pool water temperature cannot be maintained below 110°F.
(1.5 points)

b(1) Five (5) or more adjacent control rods not inserted below the 06 position.
(0.5 points)

b(2) Thirty (30) or more total control rods not inserted below the 06 position.
(0.5 points)

Comment:

- Response blanks should occur at or near the end of a test item after sufficient information has been provided to establish exactly what the problem is and what frame of reference should be used to arrive at the solution.
- Response blank labels are unnecessarily complicated b(1), b(2), (a.1), (a.2), versus 1, 2, 3, and 4.
- If blanks are used, they should either be of sufficient size to write in the answers, or a separate location should be provided for the answers.
- There is no apparent reason for assigning different values to the four responses, and the trainee will not be aware of the differential weighing.

TEST ITEM EXAMPLES

C.2 Matching Test Item Examples

Satisfactory:

Test item:

Select the answers needed to complete each sentence. Write the letter of each answer in the blank provided.

In the containment air return system, both fans are actuated upon the _____ actuation signal but are delayed starting for _____ minutes. They continuously draw air from the dome of the containment vessel and from the following pocketed spaces _____, _____, _____, and _____. (6 points)

- | | |
|------------------------------------|-----------------------------------|
| (a) High pressurizer pressure | (g) Accumulator spaces |
| (b) Containment high-high pressure | (h) Lower containment compartment |
| (c) Two | (i) Cable penetration area |
| (d) 10 | (j) Steam generator enclosures |
| (e) 60 | (k) Pressurizer enclosure |
| (f) Instrument room | (l) Upper containment. |

Answer:

(b), (d), (f), (g), (i), and (k).

TEST ITEM EXAMPLES

Unsatisfactory:

Test item:

In the containment air return system, both fans are actuated upon a _____ actuation signal but are delayed starting for _____ minutes. They continuously draw air from the dome of the containment vessel and from the following pocketed spaces _____, _____, _____, and _____. (6 points)

- | | |
|-----------------------------------|------------------------------------|
| (a) One hour | (g) Ten minutes |
| (b) Instrument room | (h) Containment high-high pressure |
| (c) HPZ | (i) Cable penetration area |
| (d) Accumulator spaces | (j) S/G enclosures |
| (e) One minute | (k) Pressurizer enclosure |
| (f) Lower containment compartment | (l) Upper containment compartment |

Answer:

(h), (g), (b), (d), (j), and (k).

Comment:

- Directions are implied but not given.
- Abbreviations are used when complete terms should be specified. (HPZ, S/G).
- Distractors are not logically grouped (i.e., by signal, time, and spaces).

TEST ITEM EXAMPLES

- Answers are not appropriate for blanks (i.e., ten minutes).
- Number of points available has not been stated.
- Blanks are not same size.
- Avoid specific determiners before blanks, like "a" or "an."

Test item:

Subatomic particles have certain properties. Match the names of the following particles with their properties by writing the letter representing the particle next to its properties: (10 points)

<u>PARTICLES</u>	<u>PROPERTIES</u>
(a)—Alpha	___1. Very small mass, one negative charge
(b)—Electron	___2. Very small mass, one positive charge
(c)—Meson	___3. Mass 1 unit, no charge
(d)—Neutrino	___4. Mass 1 unit, one positive charge
(e)—Neutron	___5. Mass 4 units, two positive charges.
(f)—Positron	
(g)—Proton	

Answer:

1 (b), 2 (f), 3 (e), 4 (g), and 5 (a) (2 points for each answer)

TEST ITEM EXAMPLES

C.3 Multiple-Choice Test Item Examples

Satisfactory:

Test item:

After a long refueling outage, a reactor startup is performed, and criticality is achieved in the source range. How would you hold reactor power level constant (in the source range) for 8 hours while surveillances are conducted? (2 points)

- (a) Adjust rods
- (b) Adjust boron concentration
- (c) Adjust reactor coolant system temperature
- (d) No adjustments required.

Answer: (a).

Unsatisfactory:

Test item:

During a cold S/U, a long refueling outage, with power level below the point of adding heat, you are asked to hold the reactor power level steady for several hours while surveillances are being conducted. During this period of time you may be doing any of the following to hold power constant. Pick the best answers:

- (a) Withdrawing rod slowly
- (b) Inserting rods slowly
- (c) Diluting boron slowly

TEST ITEM EXAMPLES

(d) Borating slowly

(e) Holding rods and boron at a constant level.

Explain your answer.

Answer: (a) and (b).

Comment:

- The stem suggests more than one response.
- There should be one clear best or correct response.
- An essay test item is mixed with a multiple-choice test item.
- The conditions in the stem are not precise; confusion exists as to the best answers.
- Acronyms such as S/U should be avoided.
- Available points are not stated.

Test item:

Why are reactivity adjustments required to hold the reactor power level constant (in the source range)? (2 points)

- (a) To compensate for doppler feedback
- (b) To compensate for the condition that K_{eff} cannot be exactly one
- (c) To compensate for power coefficients

TEST ITEM EXAMPLES

- (d) No adjustments required, the inherent power coefficients will maintain the reactor at constant power.

Answer: (b).

C.4 Essay Test Item Examples

Satisfactory:

Test item:

Assume the reactor is at 100% power. For each of the following reactor protective system trips, state (a) the sensed parameter, (b) the limiting safety system setting (LSSS), and (c) the protection provided.

- (1) High power level trip (3 points)
- (2) Low reactor coolant flow trip (3 points)
- (3) Steam generator differential pressure trip (3 points)

Answer:

- (1) (a) Neutron flux (thermal backup) is sensed.
(b) LSSS @ 100% is 107% of rated power.
(c) Prevents clad damage due to reactivity excursions too rapid to be detected by pressure/temperature trips.

TEST ITEM EXAMPLES

- (2) (a) Reactor coolant system (RCS) differential pressure across each steam generator (SG) is sensed.
- (b) LSSS @ 100% is 95% of max flow.
- (c) Protects the core against departure from nucleate boiling (DNB).
- (3) (a) Secondary differential pressure across each SG is sensed.
- (b) LSSS @ 100% is 135 psid.
- (c) Protects against departure from nucleate boiling (DNB) and high linear heat rate.

Unsatisfactory:

Test item:

For each of the reactor protective system trips listed below, state the sensed parameter, the limiting safety system setting, and what each trip protects against. Only the LSSS that would apply at 100% power is required. (3.5 points)

- (a) High power level (1.0)
- (b) Low reactor coolant flow (1.0)
- (c) Steam generator differential pressure (1.0)
- (d) Containment high pressure (sensed parameter is NOT required). (0.5)

TEST ITEM EXAMPLES

Answer:

- (a) Neutron flux (thermal backup) is sensed (0.3)

LSSS @ 100% is 107% of rated power (0.2)

Prevents clad damage due to reactivity excursions too rapid to be detected by pressure/temperature trips (0.5)

- (b) RCS differential pressure across each SG is sensed (0.3)

LSSS @ 100% is 95% of max flow (0.2)

Protects against DNB (0.5)

- (c) Secondary differential pressure across each SG is sensed (0.3)

LSSS @ 100% 135 psid (0.2)

Protects the core against DNB (0.5)

- (d) LSSS @ 100% is 5 psig (0.1)

Assures reactor is tripped whenever SG is sensed (0.4)

Reference: FCS TS 1.3, p. 1-7 to 1-10.

Comment:

- The developer has assigned different values to responses (i.e., state the sensed parameter, limiting safety system setting, and what each trip protects against).
- Abbreviations are used instead of complete wording.

APPENDIX D
EXAM TEST ITEM REVIEW CHECKLIST

INTENTIONALLY BLANK

EXAM TEST ITEM REVIEW CHECKLIST

The following checklist presents points that should be incorporated when constructing test items from the learning objectives for the written examination.

D.1 General Guidance

- Does the concept being measured have a direct, important relationship to the ability to perform the job?
- Does the test item match the learning objective? Does each test item measure one specific concept and reflect the actions, conditions, and standards of the objective being tested?
- Is the test item clear, concise, and easy to read? Does the test item contain only information relevant to the problem posed? Could it be stated more simply and still provide the necessary information? Can it be reworded or split up into more than one test item?
- Does the test item provide all necessary information, conditions, and assumptions needed for a fully correct response?
- Is the test item written at the highest appropriate level of knowledge or ability for the job position of the trainee being tested?
- Is the test item grammatically correct?
- Is the test item free of tricky wording and clues to the correct answer?
- Is the test item free of unnecessary difficulty or irrelevancy?
- Is the reading level appropriate for the target audience?

EXAM TEST ITEM REVIEW CHECKLIST

- Is the test item limited to one concept or topic?
- Are directions completing individual test items necessary for clarity? Do they repeat standard instructions provided to trainees at outset of testing?
- Does the test item have face validity?
- Are key points underlined?
- Is each test item separate and independent of all other test items?
- Are all parts of the test item on the same page?
- Are all required material, drawings, and accompanying test items clearly identified?
Can the trainee easily locate them?
- Is there sufficient space provided for answers?
- Are the less difficult test items at the beginning of each section?
- Have your test items been reviewed by others?

EXAM TEST ITEM REVIEW CHECKLIST

D.2 Point Values

- Point values are specified for whole test items and all subordinate parts (if more than one response is required).
- Values are assigned relative to other test items in terms of:
 - significance of successful performance of associated objective to task performance
 - learning difficulty and cognitive level
 - number of responses required
 - difficulty of problem.
- Specific values are assigned for parts of short-answer and essay test items.
- Test items which test the same objective have comparable point values.

D.3 Short-Answer Test Items

- Is there one, short, definitely correct answer for each test item?
- Does the scoring key follow directly from the test item?
- Are clues to the answer avoided?
- Is the required degree of precision specified? For test items requiring computation, specify the degree of precision expected. Try to make the answer turn out to be whole numbers.
- Are the test item statements simple and direct without extensive qualification?

EXAM TEST ITEM REVIEW CHECKLIST

- Are the test item blanks the same length regardless of the number of words to be entered?
- Does the test item wording avoid grammatical clues to the correct response such as "a" or "an" before the blank?
- Limit the space allocated for each answer to encourage a single word or short phrase.
- For fill-in-the-blank test items, arrange the blanks to be of equal but adequate length.
- For fill-in-the-blank test items, do not omit words that are interdependent.
- For fill-in-the-blank test items, do not make sentences unrecognizable by leaving too many blanks.
- For a completion table, do not have more than six items in one column.
- For a completion diagram, include only necessary features and ensure that components are clearly referenced.
- Compose a detailed model answer, identifying important statements, steps, or parts, and allocate points for each test item subpart.

D.4 Multiple-Choice Test Items

- Does the test item have one focused topic, making it something other than a collection of true/false test items? Does the stem clearly express a single problem in a direct manner followed by response?
- Does the test item provide sufficient information to provide a basis for formulating the response?

- Is the test item or problem defined in the stem?
- Are tricky or irrelevant test items avoided?
- Are the answer options homogeneous, highly plausible, and comparable length?
- Are "none of the above" and "all of the above" avoided?
- Is one response clearly the correct or best answer accompanied by three or four distractors?
- Is each test item stated positively, unless the intent is to test knowledge of what not to do?
- Is the test item free of "specific determiners" (e.g., logical or grammatical inconsistencies, incorrect answers which are consistently different, verbal associations between the stem and the answer options)?
- Are common misconceptions used as distractors?
- Are the answer options of the test items ordered sequentially?
- Is the test item free of trivial distractors? Does the stem include irrelevant, trivial detail or instructive information?
- Are the references, attachments, and data included in the stem or identified following the stem? The drawings accompanying the stem may proceed, be to the right of, or below responses or distractors as space permits.
- Are the correct responses varied from test item to test item in a random fashion?
- Are there equivalent and/or synonymous options that rule out both options for a trainee who recognizes the equivalence.

EXAM TEST ITEM REVIEW CHECKLIST

D.5 True or False (Multiple-Choice) Test Items

- Does the test item address only one idea or concept in a declarative statement?
- Is the test item either completely true or completely false?
- Is the test item too long or overly complex with qualifying statements?
- Is the test item free of absolute and indefinite terms such as always and sometimes and does it contain no double negatives?
- Are false statements plausible and not false because of trivial detail?

D.6 Matching Test Items

- Are tricky or irrelevant test items avoided?
- Is there a clearly correct answer or answers to the test item?
- Are clues to the answer avoided (e.g., grammatical clues, response patterns)?
- Do the directions clearly tell the trainees the basis on which to make the match and how to indicate their answers?
- Do the directions tell whether responses can be used more than once?
- Is each response a plausible answer for each premise?
- Are there more responses than premises if each response can only be used once?
- Are the responses arranged on one page in a logical order?

EXAM TEST ITEM REVIEW CHECKLIST

- Is the test item arranged so that the trainees can mark their answers easily?

D.7 Essay-Format Test Items

- Is the test item clearly and concisely worded WITHOUT AMBIGUITY?
- Does the test item illicit the correct response and no other responses? Does it clearly place bounds on the required response?
- Are point values, expected time to respond, expectations for exact answers or estimates clearly identified?
- Is enough information supplied to allow the knowledgeable trainee to correctly respond?
- Is the test item free of negative statements?
- Is the sample answer constructed to minimize subjectivity in grading? Are all significant requirements clearly identified, each with specific point values? Are alternate acceptable responses also included with the sample answer, where appropriate?

D.8 Drawing or Labeling Test Items

- Are clear instructions provided concerning what is to be drawn, labeled, or sketched?
- Are clean, readable illustrations provided with parts to be labeled specifically identified?
- Does the test item require only one type of response per diagram?

EXAM TEST ITEM REVIEW CHECKLIST

- Are spaces provided for answers that require labeling a given drawing?
- Does the test item provide sufficient space for the required drawing/sketch?
- Are the points allocated for all parts of the drawing/sketch?

CONCLUDING MATERIAL

Review Activity:

<u>DOE</u>	<u>Operations Offices</u>
AD	AL
DP	CH
EH	FN
EM	ID
ER	NV
FM	OR
NE	RF
	RL
	SAN
	SR

Preparing Activity:

DOE-EH-31

Project Number:

6910-0064

National Laboratories

- ANL
- BNL
- FNAL
- INEL
- LANL
- LLNL
- MNREL
- ORNL
- PNNL
- SNL