Test Item Writing Manual
May 2000
Format of Test Questions

It is important to have well developed test content specifications in order to develop valid tests. However, the best developed test content specifications will not ensure the development of a valid and reliable test if certain basic rules are violated when writing test items (questions). The writing of test questions is the most important and time consuming step in test development. When writing test items to fit test content specifications, there are a variety of test question format types from which to choose. The most frequently used formats include: (1) true-false, (2) matching, (3) short answer essay, and (4) multiple choice. Of these four, the multiple choice format is the only format that is acceptable for CWEA certification examinations.

The Multiple Choice Format

The multiple choice format is considered the most effective for developing standardized tests. Multiple choice items with four alternatives (i.e. three distractors and one correct answer) have been found to work best for developing standardized tests. Advantages of the multiple choice format include:

- **Versatility of Measurement.** Multiple choice items may be developed to measure higher mental processes (i.e. application, synthesis, analysis, and evaluation) as well as simpler levels of understanding.

- **Range of Content Measured.** Given that many multiple choice items may be answered in a given period, a wide range of content areas may be measured by one test.

- **Reduced Guessing.** For example, in a 10 item test with four choices per item (i.e. a, b, c, d), the probability of obtaining a score of 7 by chance (guessing) is 1 in 1000.

A good multiple choice item consists of three well developed item components:

1. Question, or item premise (i.e. the part preceding the choices)
2. Correct answer
3. Distractors (i.e. incorrect choices)

**Question or Premise.** The question or premise of the item should be a complete expression of a problem, should avoid excess verbiage, ambiguities and negatives which tend to obscure the point, and should contain only qualifying information that applies all of the alternative choices. There are two correct and appropriate forms of the multiple choice question: (1) the direct question, and (2) the incomplete statement. Each of these formats are shown below:

**Direct Question Example:**

*Which of the following tools should be used to cut PVC pipe?*

**Incomplete Statement Example:**

*When cutting PVC pipe you should use a _________.*
Beyond the direct question or incomplete statement formats, multiple choice questions can have direct answers, or can have qualified answers (both of these are acceptable). For example:

<table>
<thead>
<tr>
<th>Absolute Question Example:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which of the following tools should be used to cut PVC pipe?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Qualified Question Format:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which of the following tools is best for cutting PVC pipe?</td>
</tr>
</tbody>
</table>

**Correct Answer.** The correct answer should be a grammatically correct extension of the question or premise, should be the only correct answer or definitely the best answer, and should not contain accidental clues that set it apart from distractors.

**Distractors.** Distractors are the wrong answers. The purpose of distractors is to attract test takers who do not know the correct answer to a question. Only when distractors are equally attractive will a multiple choice item be performing at its best. That is, each of the three distractors should be equally plausible to a poorly qualified examinee. Distractors should grammatically correct extensions of the question or premise, should definitely be inferior or less correct than the correct answer, and should appear distinct from one another and plausible to those test takers who do not know the correct answer.

Correct answer and distractor formats that are acceptable include: (1) grammatically correct extensions of the question, (2) options that are independent, and (3) options that are similar in format, complexity, length, etc.

**Sources of Relevant and Irrelevant Variance in a Test Question**

Individual test questions, and whole tests, are designed to discriminate. The key is to get them to discriminate among examinees on the right factors: the knowledge, skill and ability, as needed to perform essential duties of the occupation. However, test items and whole tests will discriminate among people even if poorly written or if they are not adequate measures of knowledge, skill or ability as intended. For example, if a question or test is carefully written then we (CWEA and you) can infer from the test results who has the knowledge, skills and abilities to be certified at a given grade level. A poorly written question or test will still discriminate, but if knowledge, skill or ability are not being measured, then unintended variables will arise to take their place. These include such candidate factors as ethnicity, age, cultural background, or other unintended cognitive factors such as reading comprehension, intelligence, aptitude for guessing, or aptitude for something other than what is supposed to be measured. Therefore, relevant variance (discrimination among examinees) includes knowledge, skills and abilities needed to perform the essential duties; these are well outlined in the test content specifications.

Irrelevant variance in the outcome of the test (i.e. errors) will be generated when you write test questions that have any of the following characteristics:

- **Complexity is too high.** For example, don’t ask compound questions. Ask straight forward questions.

- **Reading level is too high.** Reading level is a function of word complexity and sentence length. Ask your questions in the most direct manner possible.

- **Grammar is incorrect or awkward.** Make sure that each of the four options, including the correct answer, are grammatically correct extensions of the question, and that each can stand alone as a statement.
• **Punctuation is inconsistent, missing, or wrong.** Punctuation for a question and each of the possible responses is important. Use clear and correct punctuation so the examinees are not mislead.

When a candidate misses a compound question, or one with an artificially high reading level, we don’t know if s/he missed it because we confused him/her, or s/he really doesn’t have the knowledge we are trying to measure. The errors listed above will dramatically increase the difficulty level of a question, and can lead to two type of errors: (1) Failure of an otherwise qualified candidate; or (2) passing of an otherwise incompetent candidate. Accurate tests, that focus on the “relevant sources of variance” will minimize these two errors, and will reduce the possibility of discriminating against examinees.

**Rules for Writing Test Questions**

The following “rules” or guidelines will help you write multiple choice questions, or select good multiple choice questions. By following these rules, you will find that item writing is not difficult.

1. Avoid the sources of irrelevant variance described above (e.g. complexity too high, reading level too high, grammar or punctuation errors).

2. Each question must be independent of all other questions. Do not ask a question that relies on a correct response to a different question.

3. The question or premise should present a single and clearly formulated problem. Do not ask compound questions.

4. The question should not contain extraneous words. Keep it simple.

5. Include in the premise or question words or phrases that are repeated in the alternatives.

6. Alternatives should be homogenous.

7. Alternatives should not overlap. Synonymous responses should be eliminated.

8. Avoid negatives. Never use any of the following question formats:

   All of the following are correct, except:
   Which of the following is not correct?

9. Never use any of the following response options:

   All of the above.
   None of the above.
   Both a. and c.
   Neither b. nor d.

**Test Item Development Form**

The Test Item Development Form is used to write new multiple choice test questions. This form contains the following elements:

**Identification of the occupational family (vocation).** Enter the name of the occupational family (vocation) in this space.
Identification of the grade level for which the question will apply. Enter grade level, 1 through 4, that this test item applies to in this space. The operational definition of the grade level is provided on each test content specification in a chart prior to the listing of essential duties. The chart is provided to ensure that the complexity level of an item is written at the appropriate grade level for which a knowledge, skill, or ability is being tested. For example, a specific knowledge, skill, or ability is required at each grade level within a vocation; however, the knowledge, skill, or ability is required at a different level of complexity. At Grade 1, simple recall and comprehension is all that is needed of the knowledge, skills, or ability; however, at Grade 2, application of the knowledge, skill, or ability is needed. For each vocation, the complexity level of the required knowledge, skills, and abilities should vary according to grade level.

The question or premise. Enter the exact wording of the question or premise in this space.

The correct answer. Enter the correct answer to the multiple choice question in this space.

Three distractors (wrong answers). Enter the three wrong answers in these spaces.

Essential duty linkage. The essential duties are listed according to a three digit reference number. The first digit identifies the grade level within the vocation and the last two digits identify the sequence in which the essential duty is listed. For example, a three digit number of 212 identifies that an essential duty is at Grade Level 2 and is the twelfth essential duty listed. You are responsible for documenting the reference number corresponding to the specific essential duty that a multiple choice item directly measures or reflects. One essential duty will be identified for each question. This is needed for validation purposes.

Knowledge, skill, or ability linkage. KSAs are listed according to a three digit reference number, similar to those used for the essential duties. All KSAs are equal in weight. That is, no single KSA is more important than any other KSA. You are responsible for documenting the reference numbers of 1 to 3 KSAs that a multiple choice item directly measures or reflects. This is needed for validation purposes.

Taxonomic index of the question. The taxonomic index is needed to document the primary level of the question in relation to the KSA that is measured. This index provides a range for documenting the level of the KSA measurement from least to most complex. The range from least to most complex includes:

1. **Knowledge.** Defined as the remembering of previously learned material and recalling appropriate information. No real understanding is implied at this level. Examinees are expected to recall or recognize information, ideas, and principles in the appropriate form in which they were learned. At the knowledge taxonomic level, the appropriate question format would include: *Who, what, when, where, how…?* Or *Describe…*

2. **Comprehension.** Refers to grasping the meaning of material and understanding the generalization of facts and how they relate to one another. Examinees are expected to translate, comprehend, or interpret information based on prior learning.

   At the comprehension taxonomic level, the appropriate question format would include: *What occurs next?* Or *Retell…*

3. **Application.** Refers to the use of learned material in new situations to demonstrate that ideas and facts are comprehended. Examinees are expected to select, transfer, and use data and principles to complete a problem or task with a minimum of direction.

   At the Application taxonomic level, the appropriate question format would include: *How is this an example of?… How is this related to?… Why is this significant?*
4. **Analysis.** Determining the structure of something by being able to break it down into its component parts. Examinees are expected to distinguish, classify, and relate the assumptions, hypotheses, evidence, or structure of a statement or question.

At the Analysis taxonomic level, the appropriate question format would include: What are the parts or features of...? Classify... according to...Outline/diagram... How does...compare/contrast with...? What evidence can you find for...?

5. **Synthesis.** Putting together facts to come up with the solution to a problem or the creation of something new. Examinees are expected to integrate and combine ideas into a product, plan or proposal that is new to him/her.

At the Synthesis taxonomic level, the appropriate question format would include: What would you predict/infer from...? What ideas can you add to...? How would you create or design...? What might happen if you combined...? What solutions would you suggest for...?

6. **Evaluation.** Making judgments about the value of materials for specified purposes. Examinees are expected to appraise, assess, or critique information on the basis of specific standards and criteria.

At the Evaluation taxonomic level, the appropriate question format would include: Do you agree that...? What do you think about...? What is the most important...? Place the following in order of priority...? How would you decide about...? What criteria would you use to assess...?

A taxonomic index for each test item is needed to ensure that the complexity of an item is written for the appropriate grade level for which a knowledge, skill, or ability is being tested.

**Difficulty level of the question.** The difficulty level of the item is an estimate of the proportion of test takers who will answer the item correctly, from those who attempt it. Difficulty levels within the range of .25, .3, .35, .40, and .45 are considered hard items. For example, an item with an estimated difficulty level of .35 means that only 35% of those who attempt the item will probably get it correct. Items with such difficulty levels should be avoided. Since the usefulness of certification examinations to make inferences about probable job performance based on test scores, including difficult items will not reveal a true indication of one’s level of competency for job performance.

Difficulty levels within the range of .50, .55, .60, .65, and .70 are considered moderately challenging items. These items are not so hard that only a small percentage of test takers may get them correct and not so easy that almost everybody will get them correct. Ideally, a multiple choice item with four responses should be written at a difficulty level that is moderately challenging to examinees.

Difficulty levels within the range of .75, .80, .85, .90, and .95 are considered easy items. A high percentage of test takers who attempt such items will answer them correctly. It is not good practice to have very easy test items on an examination.

**Identification of the author of the question.** The author of the test item completes the entire test item form, and signs this form upon completion.

Identification of the Subject matter expert validating the question. Each question that is written must be reviewed and accepted by a second subject matter expert. This individual will review the work of the author, note any differences of opinion including item edits, linkages, taxonomic
index, and difficulty level estimates. Finally, the reviewer will sign and date the form. Agreement with the author, including minor editing, validates the item.

Writing Test Questions

The following steps should be taken to write a test question:

1. Identify the essential duty area that you wish to test (these are listed on the test content specifications).

2. Identify one to three knowledge, skills, or abilities that are needed for performance of the essential duty (these are also listed on the test content specifications).

3. Formulate an idea of what you would like to ask. Write this idea in the form of a question or incomplete statement.

4. Write the correct answer for the question.

5. Write three distractors/incorrect answers for the question.

6. Review your work to make sure the question and distractors adhere to all of the “item writing rules” specified in this manual.

7. Go to the reverse side of the “Test Item Development Form” and complete all necessary documentation.

Application of Item Writing Skills

Evaluate each set of test items below and determine which is the better alternative. An answer key, with explanations, is given at the end of this section.

1. a. South America
   A. is a flat and arid country
   B. imports coffee from the U.S.
   C. has a larger population than the U.S.
   D. was settled mainly by colonists from Spain

   b. Most of South America was settled by colonists from
   A. England
   B. France
   C. Holland
   D. Spain

2. a. Which of the following states is located south of the Mason-Dixon Line?
   A. Maine
   B. New York
   C. Pennsylvania
   D. Virginia

   b. Which of the following states is not located north of the Mason-Dixon Line?
   A. Maine
   B. New York
   C. Pennsylvania
   D. Virginia
3. a. Which of the following is the best source of heat for home use?
   A. Coal
   B. Electricity
   C. Gas
   D. Oil

b. In the midwestern part of the U.S., which one of the following is the most economical source of heat for home use?
   A. Coal
   B. Electricity
   C. Gas
   D. Oil

4. a. Who discovered the North Pole?
   A. Christopher Columbus
   B. Ferdinand Magellan
   C. Robert Peary
   D. Marco Polo

b. Who discovered the North Pole?
   A. Roald Amundsen
   B. Richard Byrd
   C. Robert Peary
   D. Robert Scott

Answer Key

1. b. is the best alternative. The premise of the item should make a complete well formed problem. The first choice violates rule #3.

2. a. is the best alternative. b. contains a negative and is unnecessarily confusing.

4. b. is the best alternative. An item should contain only one correct or clearly best answer.

5. b. is the best alternative. All distractors should be plausible.