orgwide
custom
learning
solutions

Phase 2: Design
Developing Quizzes

Overview

This primer will provide you with a solid foundation on which to build your quizzes and develop item writing skills. Practical and specific item writing steps will be described, industry-specific conventions adopted by practiced item writers will be introduced, and punctuation and writing tips will be discussed.

Developing quizzes for training workshops is part science and part art. The "science" part of the exercise is the practice of incorporating proven educational psychology concepts into the design and presentation of items which will be used to evaluate an examinee's knowledge base. The "artful" part is knowing what content to include, providing a reasoned basis on which to include it, and then crafting items to accurately test for mastery of a content area.

Each section of this primer builds on the previous section; however, any section may be studied separately to develop mastery in a unique part of the item and quiz development practice. Mastery of all of the elements of quiz development requires knowledge of all the information contained in this primer.

Throughout the primer, reference is made to ancillary documents which have been created to be used with this primer (e.g., Glossary of Item Writing Terms, Item Writing Workbook, and Item Writing Worksheet). These documents and templates are part of the overall suite of instructional design templates and may be found under separate cover.

Developing Quizzes

Introduction

Quizzes should be an integral part of any training assessment program. The diagram below shows the natural evolution of a training workshop leading to the development of quizzes to assess learning in the workshop. For example, in the normal growth of an organization, new jobs might be created which would require that training be developed to train job incumbents. Alternatively, an issue might arise or a significant cultural change might take place in the organization leading to a need for the organization to take steps to either address the issue (through training) or ameliorate and manage the cultural shift, again through training. After a general need is recognized within the operation, the typical next steps would include (1) brainstorming the training concept, and (2) conducting a formal needs analysis resulting in a refined and final training concept that can be passed to the SMEs to serve as the catalyst and basic framework for the development of a training curriculum and workshop.

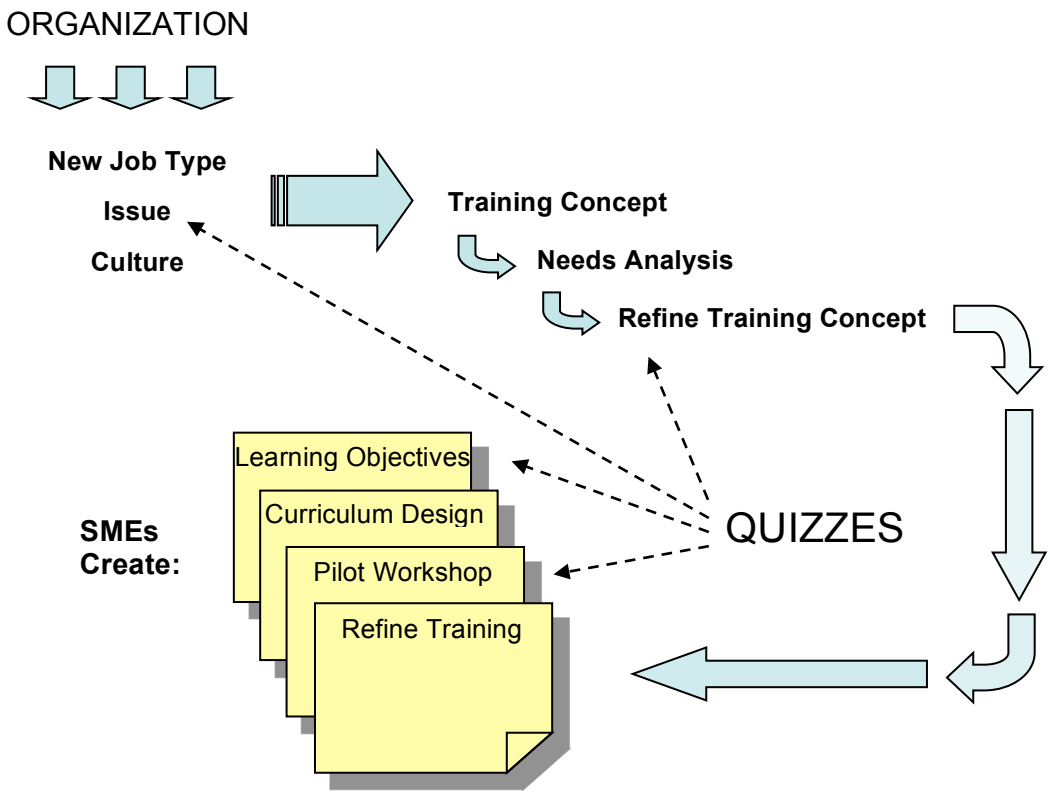


Diagram 1. LINKING QUIZZES TO NATURAL STAGES OF WORKSHOP DEVELOPMENT

The next step in an ideal instructional design methodology would be to create learning objectives. As illustrated in the diagram, quizzes should be developed from the learning objectives initially, but can also be linked to several following phases of the training design. This use of quizzes presumes the intent to assess mastery of training content. However, quizzes may also be used as a stand-alone tool for assessing a trainer's effectiveness in meeting specific training objectives. Administering quizzes provides a trainer many opportunities to support the overall training program objectives above and

beyond the material that is expected to be mastered. For example, quizzes may be used to do the following:

- create awareness of one's own behavior or attitudes towards a training program;
- personalize the training by exposing strengths and weaknesses of the participant's knowledge base;
- determine additional training and development needs;
- establish a readiness to experience more fully and personally the benefits of a training program;
- familiarize the participant with the expectations of the training;
- provide the participant with a personalized and objective report of their learning; and
- provide a starting point for additional coaching between participant-associates or instructor and participant.

Quizzes may be used in a training workshop to do far more than simply assess a participant's knowledge base. When used creatively, quizzes can serve to prepare the participant psychologically for future knowledge gain by fostering openness about assessment in general, and by creating a positive expectation for performance.

Summarizing, training may benefit from the use of well-written quizzes in the following ways:

- To encourage discussion;
- To create self awareness;
- To lower defensiveness;
- To promote teamwork;
- To identify problems;
- To improve course content; and
- Of course, to assess a knowledge base.

Developing Quiz Items

Writing quiz questions has long been regarded as more of an art than a science, but it is both. Those professionals who develop quizzes for a living will attest to the high level of creativity which is needed to produce quality quiz items. While most quiz developers claim to be true artists at their craft, the process of developing quiz items has evolved into a highly sophisticated scientific discipline. Modern-day quiz developers must: (1) adhere to complex rules and procedures to ensure that the items are measuring the appropriate aptitudes or knowledge; (2) conduct comprehensive job analyses to determine the quiz content for occupational examinations; (3) obtain knowledge and experience in developing and administrating the many different types of quiz formats (e.g., oral, written, and performance quizzes); (4) develop comprehensive review procedures to ensure the quiz items are free of bias; and (5) employ sophisticated item analysis procedures to determine if quiz items are contributing to the overall statistical soundness of the quiz and actually measure what the developer intends them to measure.

Before beginning construction of any assessment instrument, the quiz developer should establish the operational justifications for developing the assessment instrument: the quiz developer should determine the WHAT, WHY, and HOW of the quiz. The answers to these questions constitute the quiz development plan.

WHAT? The first question the developer must consider is: "What specifically are we trying to measure?" Those examinee characteristics which the developer is attempting to assess are referred to as the quiz dimensions. Quizzes are used to assess a multitude of content dimensions, such as: specific knowledge, skills, and abilities; occupational aptitudes; psychological and personality constructs (e.g., extroversion and leadership); and examinee attitudes, interests, and opinions. The procedures used to determine the dimensions differ, depending on the type of quiz to be developed. For example, the dimensions for occupational examinations are usually determined through a comprehensive job analysis. The dimensions for the General Aptitude Test Battery (GATB), an aptitude test used in the U.S. Employment Service, were determined through extensive factor analyses (multivariate statistical techniques) of other existing assessment devices. Quiz dimensions for psychological and personality assessment instruments are generally cognitively pre-determined, that is, psychological scales are usually developed to measure or test various theories regarding personality or psychological disorders. Quizzes that measure whether training objectives have been met—as is the intent of the quizzes to be developed here—are typically developed directly from the course's learning objectives and knowledge of the specific training content.

WHY? The second question the developer must consider is: "Why are we measuring it?" or, alternatively, "For what purpose is the quiz intended?" The purpose of the quiz greatly influences its characteristics such as: the length of the quiz; the difficulty level of the questions; and the format of the questions. A quiz which is used for certification (i.e., an instrument which attempts to assess the examinee's mastery of a specific subject area) might be quite different from one used for occupational selection (i.e., a quiz which attempts to differentiate between good and poor candidates), or from a diagnostic quiz (i.e., a quiz which attempts to assess the examinee's relative placement on various aptitudes and/or psychological constructs). Additionally, the purpose of the quiz directly influences how it will be scored. For example, the examinees' scores might be: (1) standardized by quiz administration; (2) compared to existing norms; (3) used to rank candidates; (4) subjected to multiple mastery cutoff levels; (5) scored on a pass/fail basis; and/or (6) subjected to various combinations of the above-mentioned scoring alternatives.

HOW? The third question the quiz developer must answer is, "How do we best measure what we're trying to measure?" Determining how to measure a quiz dimension involves determining first the quiz specifications, or logistical characteristics of the quiz. Quiz specification components include: (1) the quiz format; (2) the length of the quiz (i.e., how each question or subtest will be weighted to determine the examinee's total score); (3) the number of items per dimension and the number of total quiz items; (4) how each question will be scored; (5) the amount of time which will be allotted to examinees to complete the entire quiz and specific questions; (6) the technique used to sample or select items for the quiz; (7) the difficulty level of the questions; and (8) determining who will develop the questions. The developer should make decisions regarding all of the above-listed specifications prior to beginning development of quiz items.

Types of Quiz Questions: "Limited" vs. "Unlimited Response"

There are many different types of written quiz items which are currently being used in training-workshop quizzes, job knowledge tests and aptitude tests. These items can be classified as either limited response or unlimited response items.

In the limited response type, the examinees are provided with a finite number of possible answers to the examination question. Unlimited response items are typically used to assess examinees' knowledge of specific subject areas and basic mastery of various aptitudes such as one might find in clinical psychology or law. Types of limited response questions include True-False, Matching, and Multiple Choice questions.

In the unlimited response items, the examinees are not provided with any possible answers and are relatively unrestricted in determining the answer which they believe is the most correct. Types of unlimited response questions include Essay questions, Work Sample problems, and Fill-In-The-Blank questions. Unlimited response items are generally used to assess the examinees' cognitive processes and/or their comprehensive knowledge of a given subject matter. Often, the emphasis on scoring unlimited response items is not necessarily whether the examinee answered the question correctly (these types of problems often have more than one right answer), but on how the examinees arrived at their responses.

The major differences between limited and unlimited response items concern the scoring of the examinees' responses. Because of the infinite number of possible responses, the unlimited response items have to be hand-scored. This makes it extremely expensive and time consuming to score a large number of examinations. Additionally, unless a very detailed and comprehensive answer key has been developed for the unlimited response items, the assessors' ratings to the examinees' responses can be very subjective. Statistically, unlimited response items generally exhibit lower reliability, and the infinite number of possible responses makes it difficult to generate any accumulative item analysis statistics on them. Because of the limited ability to score these types of questions with the proposed LMS, they will not be discussed in this primer.

Limited response items are much more advantageous when administering an examination to a large number of examinees. Some of the advantages that they have over unlimited response items are that they: (1) can be objectively scored, which generally results in an increase in the items' reliability coefficients; (2) can be machine-scored, making them less expensive and time consuming to score, allowing for direct data entry of examination responses into a computer; (3) require less time for the examinee to respond, thus allowing for greater sampling of the different content areas; and (4) lend themselves to more sophisticated statistical item analysis and quiz analysis procedures.

Characteristics of Limited Response Types

True-False Questions. True-False (T/F) items are generally used to assess the examinee's basic understanding of a concept or subject matter domain. The typical T/F item is formatted as a statement, whereby the examinee must simply determine whether the statement is either correct or incorrect. Some advantages of using T/F items are that

they are generally easy to write; can be both objectively-scored and machine-scored; and most examinees can answer many T/F items in a short amount of time, thus, allowing for greater sampling of the subject matter.

Statistically, T/F items generally demonstrate moderate reliability. However, T/F items do not discriminate between good and poor examinees as well as other types of limited response items. This is due to the high chance correct-score (50 percent) associated with quiz items with only two responses from which to choose. If the quiz developer is planning to use a subtest composed entirely of T/F items, it is considered a good practice to include a sufficient number of questions to ensure the subtest exhibits an adequate degree of reliability.

Matching Questions. Matching-type questions are used to assess the examinees' ability to make associations between two sets of stimuli. Many types of stimuli can be effectively used in matching questions, ranging from simple-form-perception questions to complex-cognitive-analogy questions. In most cases, writing large quantities of matching items is relatively quick and easy because the quiz developer does not have to develop a number of incorrect responses for each examination question. Additional advantages to using matching questions are that they are objectively scored, and can exhibit a low-chance-correct score (depending on the number of stimuli in each of the sets).

One of the major drawbacks associated with most types of matching items is that the items demonstrate a degree of interdependence: often by establishing the correct answer for a particular item, examinees can eliminate that alternative as a possible correct response for subsequent items. In the same respect, an examinee's incorrect responses on one item could lead to incorrect responses on subsequent items. To minimize the interdependence associated with matching-type items, it is considered a good practice to include more items in the second set of stimuli than in the first. In spite of their interdependent nature, matching items generally demonstrate adequate reliability, primarily because of the large number of alternative responses for each item.

Multiple Choice Questions. The Multiple Choice (M/C) format is the most frequently used item format in educational and occupational testing. The typical M/C item consists of a problem, question, or statement, followed by three to five alternative choices, whereby the examinee is instructed to select the best, or "most correct," answer. This type of M/C format is generally used to assess the examinee's knowledge or ability in a specific content area.

Some of the advantages of M/C items are that they are objective, can easily be machine-scored, and exhibit a low—or predictable—chance-correct score for items with at least four alternatives. The M/C format is very versatile, and easily adapted to measuring complex cognitive processes, such as the ability to: classify information; to make associations/correlations between concepts; to distinguish differences in concepts or objects; etc. Of all of the limited response-type questions, the M/C format is the most reliable and is the best at discriminating between good and poor examinees. M/C items also possess more statistical power than true-false or matching items; consequently their format allows for more precise statistical analysis.

The major drawback in using the M/C format is that the quiz items can be very difficult and time consuming to develop. It is not easy to write M/C items which have only one correct answer and three plausible, yet distinctly incorrect, distractors. In addition, the quiz developer must be cautious that the M/C items do not contain any clues or errors which may enable those examinees with only limited knowledge or ability to select the correct answer. Most inadvertent clues or errors can be eliminated by requiring that all prospective quiz developers receive some form of training in writing multiple choice questions.

Anatomy of a Multiple Choice Quiz Question

Because multiple choice (M/C) questions are used so frequently in assessments, special attention is given here to discussing the vernacular used to describe M/C questions as well as how best to develop them.

Quiz developers generally refer to examination M/C questions as "items." The typical M/C item is divided into two main parts. The part of the item which asks the question or poses the problem is referred to as the stem. The second part of the quiz item contains a number of possible responses to the stem. These responses are referred to as the alternatives. The alternative which represents the correct response is referred to as the key. The remaining incorrect alternatives are called distractors.

The Stem. The stem, or premise, asks the question or gives some information which requires an answer. The stem in most multiple choice items is generally stated as a question, or as a fragment of a statement which is completed by the alternatives. The stem should state the entire problem or question, and should include all of the information necessary for the examinee to determine the correct answer. The question or problem should be understandable without having to refer to the alternatives.

The stem should be stated clearly and concisely in language which is appropriate for the aptitude or knowledge being assessed. For example, verbiage or overly difficult vocabulary in a mathematical word problem could reduce the effectiveness and quality of the item by placing too much emphasis on the examinees' reading comprehension ability instead of on the examinees' mathematical ability. It is best not to repeat information in each of the alternatives which could easily be incorporated into the stem. In a speeded quiz, verbiage could slow down the examinees and reduce the item's ability to differentiate between good and poor examinees. Statements such as "of the following ..." can be eliminated from the stem of most M/C items because: (1) they add virtually no useful information to the examinees' comprehension of the problem, and (2) the M/C format already implies that the correct response is limited to one of the alternatives presented.

While wording must be concise, the information given in the stem must also be complete enough to make only one of the alternatives correct. A good way to check the conciseness of an item's stem is to attempt to answer the question without referring to the alternatives. If multiple responses can be developed which correctly answer the question, then the stem should probably be reworded to make the question more specific. Avoid writing stems which rely on the alternatives as the sole source of

information in the quiz item as they are often confusing to the examinees, (e.g., "Which of the following statements is true?").

Some M/C stems resemble a fill-in-the-blank format, whereas the stem is composed of a statement with a "blank" space in the body of the text in which the alternatives represent the content of the statement which is missing. This type of format can be confusing to the examinees because it is often difficult to determine what the item is asking without referring to the alternatives. For this reason, the quiz developer should never begin a stem with a blank space. Whenever possible, it is best to reformat the stem so that the blank space appears at the end of the statement, i.e., reformat the stem into a "complete the statement" format.

The Key. The key should represent a correct and logical conclusion to the stem. The key should not be significantly different from the distractors, i.e., the examinee should not be able to pick out the correct answer because it is different from the other alternatives based on its length, content, wording, grammar, or reading level.

There should only be one correct answer for each M/C item. Many job knowledge-type quizzes instruct the examinee to choose the alternative which is most correct, thus allowing the inclusion of distractors which are only partially correct. Although this procedure makes it easier to develop M/C questions, the overall effect of this practice is that the item begins to lose its ability to discriminate between good and poor examinees. This is because the good examinees are just as likely as the poor examinees to choose a partially correct distractor.

All of the alternatives, including the key, should be expressed in the same units of measurement (e.g., inches, feet, miles, ounces, pounds, hours, or days, etc.). The inclusion of different units in the alternatives will require the examinee to perform conversions between the different units so as to compare the relevance of the alternatives. If an item is intended to assess the examinees' knowledge and/or ability to perform unit conversions, it is best to express the conversion in the item's stem, thus forcing the examinee to perform the same conversion on all of the alternatives.

The Distractors. The purpose of the distractors is to lure the less able examinees away from choosing the correct answer. The ideal distractors are those wrong answers which are frequently chosen by the less able examinees. (This can be ascertained by conducting a statistical item analysis.) A distractor is generally classified as poor if too many or too few of the examinees are choosing it. Poor distractors often result in a decrease in the item's difficulty, reliability, face validity, and the ability of the item to differentiate between knowledgeable and clueless examinees.

Distractors are usually the most difficult part of an examination item to develop. While a distractor must be distinctly incorrect, it should also represent a plausible alternative to the average quiz examinee. Most distractors can be classified as either: (1) a true statement which does not satisfy the requirements of the stem, (2) a true statement which is either too broad or narrow in content to accurately complete the stem, or (3) a false statement which appears plausible to the less able examinee. One approach to writing effective distractors is to replicate the common errors, misconceptions, illogical conclusions, or erroneous ideas that might be applied by the less able examinees. For

example, a possible distractor for a mathematical addition item might be developed by omitting the carrying over of a number.

Distractors should be parallel in form, content, length, and difficulty to the keyed response and to the other distractors. The degree of similarity among the alternatives determines the difficulty of the item. Strongly dissimilar alternatives may require only minimal knowledge or ability to ascertain the correct answer. Homogeneous alternatives require more detailed knowledge or ability in order to distinguish the correct response from the distractors. Additionally, the quiz developer should not develop distractors which are the direct opposite of the keyed response. Such distractors might be a clue to quiz-wise examinees; if they know that the distractor is definitely incorrect, deductive reasoning might lead them to correctly conclude that the direct opposite of that false statement is, in fact, true.

Avoiding Item-Writing Clues and Errors

M/C items which contain clues or errors may enable an examinee with only limited knowledge or ability to answer the item correctly. Inadvertent clues in an item may help examinees: (1) determine the correct answer, or (2) eliminate one or more of the distractors, thus, increasing their chances of choosing (or guessing) the correct answer. In addition, item-writing errors (such as grammatical errors) in the key could inadvertently distract quiz-wise examinees away from choosing the correct answer. There are many different types of item-writing clues and errors which decrease the ability of a multiple choice item to differentiate between good and poor examinees; some of the more common errors are described below.

Grammatical Errors. Grammatical errors occur when one or more of the alternatives do not grammatically follow the premise. Some of the more common types of grammatical errors include: (1) a stem which contains a plural pronoun, while an alternative contains a singular object, and vice versa; (2) misuse of the indefinite articles "a" and "an," (e.g., when the stem, which needs to be completed by the alternatives, ends in the word "an" and one of the alternatives begins with a consonant); and (3) errors in subject-verb agreement between the stem and an alternative.

Vocabulary Clues. Vocabulary clues are words which appear in the stem and/or the key which assist the examinee in choosing the correct answer. There are four basic types of vocabulary clues: (1) the inclusion of the same word(s) in the stem as in the key; (2) the inclusion of words with the same meaning in both the stem and the key; (3) the inclusion of words which have the same prefix in the stem and in the key, such as "incapable" and "incompetent"; and (4) the inclusion of words in the stem and in the key which are commonly associated with each other, such as "gravity" and "force."

Length of Alternatives. It is important that all of the alternatives be approximately the same length. When one alternative is distinctly longer than the other alternatives, quiz-wise examinees often choose the longer alternative recognizing that it, more than likely, represents the correct answer. Research in item writing and quiz development has substantiated this assumption.

Similar Alternatives. Avoid writing multiple alternatives which are similar in meaning. If one of those alternatives is the key, then the item may possess two correct answers. If both alternatives are distractors, then quiz-wise examinees may be able to eliminate both alternatives based on the assumption that the question would not have two correct answers.

Key Not Equally Distributed Among Alternatives. When developing items, the quiz developer should make a conscious effort to ensure that the key is distributed approximately equally among the alternatives. A common strategy among quiz-wise examinees is "if you do not know the answer, pick alternative C" (some strategies advocate choosing alternative b). These strategies are based on studies which demonstrate a disproportionate number of correct responses for alternatives b and c in typical M/C quizzes.

Overlapping Alternatives. Overlapping alternatives are generally found in questions with alternatives which represent a numerical range of possible responses. For example, in an item which uses ranges of ages in the alternatives, the two most common overlapping errors are: (1) overlapping of the endpoints of the alternatives' ranges, e.g., "5 to 9 years," and "9 to 11 years"; and (2) one alternative's range being completely contained in another alternative's range, e.g., "11 to 15 years," and "11 to 19 years." The justification for avoiding overlapping alternatives is to ensure that there is only one correct answer for a given item.

Independence of Item Content. Avoid writing items that contain similar information such that if the two items were used together in the same quiz, one item would provide a clue to the other item. These types of errors are common to job-knowledge quizzes which use multiple items on the same subject matter.

All Distractors Differ from the Key in the Same Manner. This type of error is generally related to the temperament of the item's content, i.e., the positive or negative characteristics associated with the item's subject matter. Distractors with temperaments opposite the key are relatively easy to write and are usually distinctly incorrect (for that reason). However, when all of the distractors' temperaments differ from the key in the same manner, quiz-wise examinees will recognize this as a clue to which response represents the correct answer.

General Guidelines for Writing Items

- Write about an important on-the-job situation.
- Write the item as clearly as possible in the working language of the job.
- Have one central problem in each item that is clearly stated.
- Have one, and only one, correct answer for each item.
- Avoid writing negative questions where possible.
- Avoid writing very difficult or very easy items.
- Ensure that the distractors:
  - ✓ are plausible—but not correct;
  - ✓ fit the stem grammatically; and
  - ✓ are parallel in idea and form.
- Avoid giving a clue about the correct answer.
- Do not use "all of the above" or "none of the above" as alternatives.
- Avoid making the correct answer the longest of the alternatives.
- Arrange alternatives with imbedded numbers in ascending or descending order.

Item Development Checklist

THE STEM

- Does the stem present a single, clearly stated problem?
- Is the stem short and concise?
- Does the stem avoid the use of negative qualifiers?
- Is there any information common to all the alternatives which could be included in the stem?

THE KEY

- Is the key the best and only correct answer among the alternatives?
- Does the key represent a logical extension of the stem?
- Is the key parallel to the distractors in form, length, and content?

THE DISTRACTORS

- Does each distractor appear plausible?
- Does each distractor represent a logical extension of the stem?
- Do the distractors represent incorrect responses?
- Are each of the distractors parallel in form, length, and content?

ITEM CONTENT

- Are the alternatives arranged in a consistent, logical order?
- Is the item written at the appropriate level of difficulty?
- Does the item:
  - ✓ avoid controversial material?
  - ✓ avoid stereotypes relating to gender and ethnicity?
  - ✓ avoid the use of undefined alternatives?
  - ✓ appear practical and logical with regard to everyday situations?
  - ✓ require that the examinee know other information not contained in the item to answer it correctly?

ITEM WRITING CLUES AND ERRORS

- Do all of the alternatives grammatically follow the stem?
- Are there vocabulary clues in the stem and in the keyed response?
- Are all of the alternatives approximately the same length?
- Is the keyed response randomly and equally distributed between the alternatives?
- Does the item avoid the use of definitive and partial qualifiers?
- Does the content of the alternatives overlap?
- Do the distractors differ from the key in the same manner?
- Are any of the distractors direct opposites of the keyed response?
- Is the item's content independent from other items in the quiz?