

Guidelines for Writing Effective Tests: A Practical "Short Course" for Test Authors

Trainers, and others in organizations who author tests, often have a high degree of technical and subject matter expertise, but many haven't been exposed to the basic principles and criteria for writing tests. This *Insight* white paper offers practical guidelines on how to write effective tests. While following the guidelines will result in better tests, we suggest that it may be worthwhile to seek professional counsel when the test is particularly important, or when the test must meet legal guidelines because it is used in some fashion to make personnel decisions.

By "effective tests" we mean that they will accomplish the objectives for which they are intended. This brings up the first guideline – the test author must be clear about the test's objectives. Some tests are designed to *certify* that people have acquired, and can apply, a certain set of skills or a body of knowledge. Other tests are designed mainly to help people further *develop* by measuring skills and knowledge, and then providing specific feedback. Excluding selection tests, those are the two main objectives, and many tests are designed to do both.

Tests can be used for purposes other than certification and development (for example, evaluating the effectiveness of training programs, measuring ROI, promoting a culture of continuous learning in the organization). Regardless of the purpose, there is an increasing awareness that even small investments in testing can significantly leverage the enormous investments companies make in developmental programs (formal training, self-study and on-the-job development, managerial coaching, etc.). The purpose of this document is to help test authors capitalize on that opportunity.

Scope of this Document

The subtitle above contains the phrase "Short Course," and the treatment of the topic here is very short, indeed. The science of writing tests is complex, and Industrial-Organizational Psychologists spend years earning their doctorate degrees to understand that complexity. This document is not intended to be a treatise on the topic, but rather a way to convey the basic principles to test authors with less formal training.

The term "test" is intentionally used in this document, instead of the more generic term "assessment." There are many different kinds of assessments, ranging from simple multiple-choice tests to day-long assessment centers. The guidelines offered here pertain to what we usually think of as "knowledge tests" – tests that can be objectively scored (versus essays, job simulations, etc.). However, as we discuss below, this does not mean that such tests can only measure simple information recall. Well-constructed objective tests can measure complex abilities to apply knowledge and skills in real-world situations.

The Test Blueprint

As stated previously, a good test starts with the test author determining the objectives for the test and what will be done with the results. Based on that determination, the next step is to prepare a *test blueprint* – a game plan for the content to be included in the test, how scores will be calculated, and other related matters. A builder cannot construct a good home without a blueprint, nor can a test author construct a good test.

Conceptually, developing a test blueprint is pretty simple and involves two basic steps:

1. Map out the domain of knowledge and skills the test takers should have – what they should know or be able to do.
2. Determine the test items (types of items, how many items, difficulty level, etc.) needed to accurately measure each of the major knowledge and skill areas.

In practice, however, developing a test blueprint can seem like such a daunting challenge that many test authors don't even bother – they just start writing test items. That's a mistake because the end result is unlikely to be a tool that is helpful to either the company or the test takers.

As a side comment, we note that one of the important benefits of linking solid testing programs to developmental initiatives is that it helps ensure those initiatives have specifically defined objectives. In this sense, good testing leads to better training and development.

FIGURE 1
Test Blueprint Matrix

| Levels of Knowledge/Skill | Topics/Sections | | | | | Total |
|-----------------------------|-----------------|----|----|----|----|-------|
| | A | B | C | D | E | |
| Knowledge of terms | 2 | 2 | 0 | 2 | 1 | 7 |
| Comprehension of principles | 3 | 2 | 3 | 4 | 2 | 14 |
| Application of principles | 2 | 4 | 3 | 5 | 3 | 17 |
| Analysis of situations | 3 | 3 | 4 | 4 | 2 | 16 |
| Evaluation of solutions | 0 | 2 | 2 | 4 | 3 | 11 |
| Total: | 10 | 13 | 12 | 19 | 11 | 65 |

Okay, we've made the point that it's very important to develop a test blueprint, but it's not all that easy to do. So what's the solution? Creating a test matrix as shown in Figure 1 is a good starting point.

In this example, the domain of knowledge and skills has been divided into five topics, or sections, represented by the columns. The rows in the matrix represent different levels, ranging from understanding terms to evaluating solutions. The numbers in the cells indicate the number of test items for each topic and level. The matrix helps the test author ensure that the content of the test reflects what was covered in the training program, or what is important to the test taker's successful performance on the job. For example, in this case, topic D is most important, and merely knowing the terms is less important than the other levels.

We'll briefly mention three other points related to the matrix before moving on. First, the different levels shown in the matrix are fairly generic and will work for many tests, but not all. You should choose whatever levels make sense for your particular situation. For example, a product knowledge test for customer service reps who interact with a bank's customers might have these three levels: knowledge of terms, product features and benefits, and ability to link products to customers' needs.

Second, how many items should be included in a test? That's not an easy question to answer, and it depends on the total domain being measured, what scores will be generated, and what will be done with the scores. Here are some guidelines:

- Obviously, a summary test that measures many content areas will need more items than a test that measures only what people have learned in a two-hour training module.
- If there is a need to measure each section accurately, in addition to a total score, then more items will be needed. A rule of thumb is that each section should have at least 8 items.
- If there are serious positive or negative consequences associated with the scores (e.g., promotion, having to take the test again, having to repeat training), then more items are needed.

Third, for simplicity's sake, the remainder of this document will discuss test development in terms of measuring knowledge and skills associated with training programs. However, the same principles can be applied to tests associated with any measurement need.

Types of Test Items

At this point, you've specified the domain of knowledge and skills your test will measure in terms of topics and levels, and the number of items (data points) in each. The next step is to decide what types of items the test will contain. It is beyond the scope of this document to go into too much detail on item types, but a few of the key points are summarized in this section.

Most types include a stem where the question or problem is stated, and response options where the test taker chooses one or more answers. Exceptions to this are noted in the following discussion. Here are the primary types of test items (objective tests only):

Multiple choice – Stem followed by several response options (usually 4 or 5), only one of which is correct. The other options are called distracters. This is the most commonly used type. The advantage is that, with careful construction, this type can be used to measure knowledge at most levels. The disadvantage is that it's hard to write good distracters for levels beyond factual recall.

What is the capital of Florida?

- A. Miami
- B. Tallahassee
- C. Orlando
- D. Jacksonville

True-false – Statement which is either true or false. The advantage is that it's the most efficient way to measure a lot of content in a short period of test time. The disadvantages are that it's hard to measure higher-level knowledge areas, and guessing (50% chance of being right).

The border between the U.S. and Canada is longer than the border between the U.S. and Mexico.

- A. True
- B. False

Multiple select – Stem followed by several response options (4-10), more than one of which may be correct. The advantage is that it is an efficient way of measuring a set of facts or concepts that cluster together. The disadvantage is that this is suitable only for certain knowledge areas.

What colors are in the American flag? Mark all that are correct.

- Red
- Green
- White
- Blue
- Black

Matching – List of premises (definitions, explanations) on the left, and response options on the right. The test taker must select the best response option for each of the premises. The advantage is that it allows the comparison of related ideas or concepts. The disadvantages are that it's not suitable for measuring isolated facts and information, and scoring can be complex.

For each concept on the left, select the word from the list on the right that best matches it.

- | | |
|--|------------------|
| <input type="checkbox"/> Test predicts future performance | A. Face validity |
| <input type="checkbox"/> Test appears a reasonable measure | B. Reliability |
| <input type="checkbox"/> Re-test scores are very similar | C. Accuracy |
| <input type="checkbox"/> Low standard error | D. Validity |
| | E. Consistency |

Ranking – List of response options the test taker must put in the proper order (by sequence, importance or some other factor). The advantage is that this is perfect when knowing the correct order is important. The disadvantages are that it's not suitable for anything else, and scoring can be complex.

Put the following steps in the correct order a test author should take in writing a new test.

- Prepare test blueprint
- Determine test objectives
- Draft test items
- Evaluate items against criteria
- Perform item analysis
- Check with subject matter experts
- Select item types to be used
- Pilot the test and modify as needed

Fill in the blank – Stem consists of a statement or question, and the test taker must supply the answer. The advantages are that it minimizes guessing, and this type is easy to write. The disadvantages are that scoring can be difficult (and sometimes subjective), and this type has little advantage over well-written multiple-choice items.

The first President of the United States was _____.

Choosing Item Types

There are many different types of test items. Which one(s) should you choose? Again, there are no hard and fast rules, but here are some guidelines that might be helpful.

- Start with multiple-choice items since they are the most versatile, and then add other types as necessary.
- Be careful not to have too many item types in the same test. This makes it confusing to test takers. There's nothing wrong with having all multiple-choice items if doing so meets your measurement objectives.
- Related to the two previous points, keep things simple. (By "simple," we don't mean low difficulty level.) While training program delivery should be varied to sustain interest, this usually isn't an issue for test takers.
- Think about the specific knowledge or skill you're trying to measure, and then consider which type of test items will be the most accurate and efficient.
- The stems for multiple-choice items don't necessarily have to be short questions or problems (but write as concisely as possible!). You can create scenarios or situations that require test takers to engage in analysis and critical thinking. You can also link more than one item to a scenario.
- Add graphics, diagrams, pictures and other information to the items as appropriate. With computer-based testing, you can also include sound and/or video files.
- Multiple choice, T-F, and fill in the blank items are usually scored as either right or wrong. The scoring for other item types (multiple select, matching and ranking) can get complicated. Make sure you know how you plan to score them before administering the test.

Having extolled the advantages of multiple-choice items, and having cautioned test authors to keep things simple, we need to backtrack just a bit. Test authors should be creative in writing tests that will best accomplish their objectives. Be creative, and don't get stuck in a rut!

Steps in Developing a Test

Before getting into the specifics of writing good test items, let's first take a higher-level view of the overall steps in developing a new test. On the previous page, the "ranking" example listed the steps. (Give yourself bonus points if you tried to order them.) The list in Figure 2 shows the steps in the correct order.

FIGURE 2

Steps in Developing a Test

1. Determine test objectives
2. Prepare test blueprint
3. Select item types to be used
4. Draft test items
5. Evaluate items against criteria
6. Check with subject matter experts
7. Pilot the test and modify as needed
8. Perform item analysis

We've already discussed Steps 1-3 – that was done first so you would have the proper context for understanding the sequence of events. Steps 4 and 5 will be covered next, and then we'll briefly touch on the remaining steps. Developing a good test is clearly more than just sitting down and writing items.

General Points on Writing Effective Tests

This section and the next provide suggestions on how to move from *theory* to *application* in writing effective tests. There have been volumes written on this topic, and some of the technical and measurement issues are rather complex, so we'll only be able to hit the high points in this document.

As you author test items, keep the most important principle in mind: **Each item must accurately measure the knowledge or skill area on which the item is focused.** This is an extremely important point, and one that is often overlooked by people who provide advice to test authors. One can get so caught up in the different criteria and psychometric nuances that you lose sight of the main point – "Is this an item a competent test taker should get right, and is it one that someone less competent might get wrong?" If the answer is "no," then it's not a good item.

If the test is associated with a training program (or some other developmental activity), the content of the program should be the focus of the test. In this case, test development should be fairly straightforward. What knowledge and skills were being developed, and what test items will help measure the degree to which that occurred? In other words, in authoring a test associated with training, you should start with the content that the test takers should have acquired (i.e., the learning objectives).

Criteria for Effective Items

There are a great many rules/criteria for writing effective test items. The checklist at the end of this document summarizes the most important ones. You should be thoroughly familiar with the criteria as you draft items (Step 4), and you can use the checklist as you carefully evaluate your work afterwards (Step 5).

Take a look at the checklist now. Many of the criteria are applicable to all tests, but some apply only to certain item types. Most criteria are self-explanatory, but the comments below will help you understand the rationale for some of the criteria. (The numbers-letters correspond to those on the checklist.)

- 1-a As a test author, you may or may not have responsibility for determining the objectives of the test. If you don't, ask someone who does.
- 1-e Tests that are not associated with a particular developmental activity must still be associated with something. There's still a need to be clear on what you're trying to measure and why.
- 2-b With unsophisticated test takers, an example of a completed item, given as part of the instructions, might be appropriate for more complex item types.
- 3-a Trying to measure more than one thing with a single item is a mistake commonly made by new test authors.
- 3-b Importance is sometimes confused with item difficulty. Something could be extremely important, but if 100% of the test takers always get the item right, it's probably trivial and should be eliminated.
- 3-d The easiest way to tell if the stem is a complete thought is to cover up the response options and see if you know what you're supposed to do.
- 3-f Write each item as concisely as you can, but not to the point of sacrificing clarity.
- 3-g Use simple words, but don't shy away from technical terms if they are important for test takers to know.
- 3-h As you ensure that all response options are grammatically correct with respect to the stem, try to avoid the "a(n)," "is/are" solution. Rewrite the item so you can measure the knowledge or skill without getting hung up on the grammar.

- 3-i Writing plausible distracters is both an art and a science, and it's very hard work. An approach some authors find helpful is to first write the stem and correct response, and then think of good distracters.
- 3-j An example of response options that aren't independent is: A. <10, B. <20, C. >40, D. >50.
- 3-m In addition to one item not leading to the correct answer on another item, make sure the test takers aren't penalized on one item because they got another item wrong.
- 3-p Occasionally, statements with definitive words are actually true, and are good items. For example, the statement "all aptitude tests that are valid are also reliable" happens to be true.
- 4-e The problem with "A and B" as a response option is that it becomes more of a logic test than a knowledge or skills test. However, it is occasionally permissible as long as there aren't too many such items in the same test.
- 7-c The reason for having the shorter text on the right is that most test takers will be reading that list repeatedly to find the correct matches.
- 7-e The reason for having an unequal number of premises and response options is that it makes the item more challenging because the test taker cannot determine the last matches via the process of elimination. However, there's nothing wrong with having an equal number.

Refining the Test

The last three steps in Figure 2 deal with refining the test to make it even better. Step 6 is to check with subject matter experts. It's always a good practice to have someone else review the test before it's administered, both from a psychometric standpoint (the criteria for good items), and a technical, subject matter standpoint. The review process is also a good time to clean up typos and grammatical mistakes.

Step 7 is to pilot the test with a small sample of people who are similar to those in the actual test taker population. A focus group works well for this. Administer the test and note the completion time for each participant. This will help you judge how much time should be allotted for real test takers. Debrief the participants afterwards to solicit input on their overall reactions, as well as their reactions to any items they felt were troublesome (unclear, no right answer, etc.).

Finally, Step 8 is to perform an item analysis once you have a large enough sample of test taker results. This is the true “test of your test,” and is an excellent way to further refine it. Again, it’s beyond our scope in this document to go into much detail, but here are a few key points:

- Be careful not to draw conclusions based on small sample sizes ($N < 30$).
- Examine items where the percent correct is low (e.g., $< 65\%$), and look at the distribution of responses across the distracters. If the correct answer is B, but almost everyone who missed it picked C, the item could be confusing or ambiguous for some reason. On the other hand, it could just be a difficult item, which is fine.
- Examine items where the percent correct is very high (e.g., $> 95\%$). It could be that the item is too easy or even trivial. At the extreme, an item with 100% correct is probably not worth retaining unless you want to keep it in for reinforcement purposes.
- Identify those distracters that were picked by an extremely low percentage of test takers (e.g., 0-3%). See if you can rewrite them to be a little more plausible.
- Examine the correlations between item scores and total test score. If high-scoring people are more likely to miss a particular item, then the item is probably defective.

Besides refining your test, item analysis is extremely valuable in improving training programs and their delivery. You’ll have quantitative data on what people don’t know or understand, a solid basis for making changes to training materials, and good feedback for trainers.

Difficulty Level and Scoring

In previous sections and in the checklist, we briefly touch on the concept of item difficulty level. You should have a range of items within sections – some easier items and some harder items. The remaining question is, how difficult should the whole test be? In general, a good test should be challenging enough that few or even none of the test takers will get 100%. Why? Because you can’t tell how much *more* the test takers actually know when they get 100% – you didn’t measure that “high.” Another reason is that feedback given to test takers is much more useful when they discover areas where they can further develop.

There are exceptions to the guideline that good tests should be challenging to even the most competent test takers. For example, some training has very strictly defined and limited learning objectives. The purpose of the test may only be to certify that people have mastered some baseline knowledge or skill, and you’re not trying to do anything beyond that.

Then we come to the issue of setting a passing score – the total percent correct required for a test taker to pass the test. Before going any further, make sure you need a passing score in the first place. When tests are used for developmental purposes, it may be preferable to simply report a total score and section scores without reporting whether people have passed or failed. However, in many situations there are compelling reasons to set a passing score, and you can’t duck the issue.

The passing score is obviously dependent on the overall test difficulty level, and there are no rules of thumb such as “the passing score should never be lower than 85%.” There are scientific, statistical methods available for setting passing scores, but here’s a simple approach that is adequate for most tests:

1. Meet with a group of subject matter experts, perhaps including managers and trainers. Review the content of the test and pose the question, “What percentage of these items do you think people should get correct to be considered fully competent to perform their jobs?” Reach a consensus as a group.
2. During that same meeting (but *after* you’ve reached a consensus), examine the distribution of total scores of actual test takers and note the percentage of people who did not pass. Ask the group if the results seem logical in terms of performance on the job and business needs for competencies. Reach a consensus on an adjusted passing score, but be careful not to unduly lower your standards of excellence.

You obviously can’t do Step 2 until after you have results from actual test takers, but at some point in the future you should look at the distribution of scores and consider adjusting the passing score as appropriate.

There are many other scoring complexities you can get into (but we won’t here), including weighting items and sections, requiring that some minimum level be achieved on each section before one can pass the overall test, different passing scores for the first attempt versus a test re-take, and whether guessing should be penalized, among others. Our only advice is to keep things as simple as possible.

Feedback to Test Takers

The real value in testing is to further the learning and development process. How well that is accomplished depends on what kind of feedback test takers receive, and when they get it. To simply tell a test taker their score and whether they passed or failed misses a whole world of development opportunities. Good feedback consists of these elements:

- Total score (and sometime normative information on how the test taker compares to others)
- Section scores, so the relative strengths and weaknesses can be understood.
- Items that were missed so the test taker understands the specific knowledge or skill deficiencies
- Development suggestions for recommended actions to take based on test results (self study, training programs or materials, on-the-job actions, etc.).

Final Tips for Test Authors

Here are a few final thoughts that might be helpful:

- Make sure your tests reflect the *application* of knowledge and skills – not just information recall.
- Don't be tricky or picky with test takers. They won't appreciate it, and it won't help the business.
- Keep track of all the tests you author. Over time, you can build a valuable pool of test items.
- Look for opportunities to leverage technology in test administration, scoring and reporting, item analysis, and other areas. Once you've written an effective test, you want the testing process to be fast, painless and cost-effective. Using technology is extremely beneficial. For information on selecting a technology partner, read the *Insight* white paper titled "Choosing an Internet Assessment Vendor: A Practical Guide for Ensuring a Value-Added Partner."

Censeo Corporation is a human resource consulting firm offering consulting services and Internet-based assessment products. The company's mission is to help clients maximize their investment in human capital by collecting and then strategically using accurate and valid information about competencies and work preferences. For more information, visit www.censeocorp.com, email info-request@censeocorp.com or call 407.645.1600 x150.

Criteria Checklist for Test Authors

Test: _____

Author: _____

Date: _____

Notes: Test authors can refer to this checklist as they write items, and then later as the author and other subject matter experts review the draft. The first three sets of criteria are generally applicable to all tests, while the last six sets apply to tests containing certain item types. Refer to previous pages in this document for explanations of some of the criteria.

| Major Category/Criterion | OK? |
|---|--------------------------|
| 1. Overall test | |
| a. Objectives for the test, and how the results will be used, are clearly defined. | <input type="checkbox"/> |
| b. Test blueprint has been written, and it maps the content for the test in terms of topics and levels. | <input type="checkbox"/> |
| c. Items in the test provide adequate coverage with respect to the blueprint (important areas not overlooked; extraneous areas not included). | <input type="checkbox"/> |
| d. If section scores are reported, there are at least 8 items for each section. | <input type="checkbox"/> |
| e. Test as a whole is a good measure of the knowledge and skills the training program (or other developmental activity) was intended to enhance. | <input type="checkbox"/> |
| 2. Instructions | |
| a. Clear instructions for the overall test (e.g., how responses are recorded, penalty or not for guessing, time allowed). | <input type="checkbox"/> |
| b. Where necessary, clear instructions for certain item types (e.g., "mark all those that are correct" for multiple-select items, whether or not response options can be used more than once in a matching item). | <input type="checkbox"/> |
| c. It is clear to the test taker whether the focus is on selecting the <i>correct</i> answer or the <i>best</i> answer. | <input type="checkbox"/> |
| 3. Criteria applicable to more than one item type | |
| a. Each item is focused on only <i>one</i> thought (knowledge or skill area). | <input type="checkbox"/> |
| b. Item is relevant and important – something people should know versus being able to look up; not trivial. | <input type="checkbox"/> |
| c. Difficulty level of the item is appropriate for the test takers. | <input type="checkbox"/> |
| d. Stem is a complete thought (i.e., the task is clear without having to look at the response options). | <input type="checkbox"/> |

| Major Category/Criterion | OK? |
|---|--------------------------|
| e. The bulk of the text is in the stem; the response options are as short as possible. | <input type="checkbox"/> |
| f. Syntax is clear and unambiguous; wording is concise and straightforward. | <input type="checkbox"/> |
| g. Low level of reading difficulty; unnecessarily big or unusual words are avoided. | <input type="checkbox"/> |
| h. Syntax doesn't preclude any of the response options; all are grammatically correct. | <input type="checkbox"/> |
| i. All distracters are plausible, and could be selected by less competent test takers. | <input type="checkbox"/> |
| j. Response options are independent (e.g., one isn't a subset of another, no opposites). | <input type="checkbox"/> |
| k. Response options are in a logical order (e.g., by size, alphabetical). | <input type="checkbox"/> |
| l. Correct response options are randomized with respect to position among distracters. | <input type="checkbox"/> |
| m. Items are independent of each other (e.g., content of one item doesn't help the test taker get some other item right). | <input type="checkbox"/> |
| n. There is a mix of items of different difficulty levels (some easier, some harder). | <input type="checkbox"/> |
| o. Within sections, or within items clustered by item type, the easier items come first. | <input type="checkbox"/> |
| p. Definitive terms like <i>always</i> , <i>all</i> , and <i>never</i> are avoided. | <input type="checkbox"/> |
| q. Where appropriate, items are clustered by type; don't have to frequently change types (e.g., T-F items appear together). | <input type="checkbox"/> |
| r. Correct response option and distracters are similar in terms of length, sophistication, etc. | <input type="checkbox"/> |
| s. Response options don't all repeat words that could be included in the stem. | <input type="checkbox"/> |
| t. Positive wording used when possible (e.g., use of NOT, EXCEPT, etc. minimized); negative words in all caps or underlined when they must be used. | <input type="checkbox"/> |

| Major Category/Criterion | OK? |
|--|--------------------------|
| u. No double negatives (ever). | <input type="checkbox"/> |
| v. No irrelevant clues as to the correct answer in stem or response options. | <input type="checkbox"/> |
| w. Items are not split between pages. | <input type="checkbox"/> |
| 4. Multiple choice | |
| a. Only one correct answer; all distracters are definitely wrong (or definitely not as good as the correct response). | <input type="checkbox"/> |
| b. Either 4 or 5 response options (with fewer than 4, guessing is a problem; more than 5 adds little value and increases test time). | <input type="checkbox"/> |
| c. "All of the above" is avoided, and "None of the above" is used sparingly (e.g., when something is calculated). | <input type="checkbox"/> |
| d. No two response options have the same meaning (the savvy test taker could eliminate both). | <input type="checkbox"/> |
| e. "A and B," "A, B and D" response option logic is used only when you've given up trying to write a better item. | <input type="checkbox"/> |
| 5. True-false | |
| a. Statement is worded precisely so it is unequivocally (but not obviously) true or false. | <input type="checkbox"/> |
| b. About 60% of the statements are false (to counterbalance the fact that people guessing are more likely to select true). | <input type="checkbox"/> |
| c. Vague terms that are open to interpretation (e.g., <i>frequently</i> and <i>sometimes</i>) are avoided. | <input type="checkbox"/> |
| d. True statements are no longer, no more specific, and no more correct sounding than false statements. | <input type="checkbox"/> |
| 6. Multiple select | |
| a. At least 4, but not more than 10, response options. | <input type="checkbox"/> |
| b. At least one of the response options is correct (none correct would be tricky). | <input type="checkbox"/> |

| Major Category/Criterion | OK? |
|--|--------------------------|
| 7. Matching | |
| a. All premises and response options are members of the same set (a similar category or topic). | <input type="checkbox"/> |
| b. Each premise has one, and <i>only one</i> , correct response option associated with it. | <input type="checkbox"/> |
| c. The longer text items (definitions, concepts, etc.) are premises on the left; the response options on the right are as brief as possible. | <input type="checkbox"/> |
| d. Irrelevant association cues between premises and options are eliminated. | <input type="checkbox"/> |
| e. Where appropriate, the number of premises and response options are not equal (e.g., more options than premises, or options can be used more than once). | <input type="checkbox"/> |
| 8. Ranking | |
| a. At least 4, but not more than 10, response options. | <input type="checkbox"/> |
| b. There is unequivocally a correct order to the response options. | <input type="checkbox"/> |
| 9. Fill in the blank | |
| a. Only one or two blanks for each item. | <input type="checkbox"/> |
| b. Correct responses are single words, very short phrases, or values. | <input type="checkbox"/> |
| c. A priori decisions are made about which responses will be scored as correct (e.g., misspelled words; will \$1,800, \$1800, and \$1,800.00 all be correct?). | <input type="checkbox"/> |