

ASTD TechKnowledge Conference and Exposition

Session TH202TI

February 28, 2007

9:45 – 12:00

Creating Fair Tests and Certifications

Nanette Miner, Ed.D.

By the end of today's session you will be able to -

- ✓ Explain and apply the rules that govern "fair" test questions
- ✓ Design various kinds of effective test questions (multiple choice, true/false, etc.)
- ✓ Explain when, how, and by whom a test should be administered
- ✓ Determine when performance-based and criterion-referenced testing are appropriate in the workplace



P O Box 1819
Bristol, CT 06011-1819
800-282-5474
www.trainingdr.com

Creating Fair Tests and Certifications

Creating a fair test or certification is more than asking twenty questions. Do you need to test for knowledge or skill? What is a passing grade? What if people "fail?" What are the legal ramifications of giving tests in the workplace? You **MUST** know what you're testing, why you're testing, and what you'll do with the information once you've tested. **When in doubt: DO NOT TEST.**

This document will cover these weighty considerations, as well as...

- The #1 rule for creating good, fair tests
- How to design various types of questions
- When should a test or certification be administered?
- What is performance-based, and how do you test for it?
- What is criterion referenced; and how critical is it to designing effective tests?

To paraphrase Odin Westgaard, the guru of test creation:

***Your test should be easy
for those who understand what you've taught,
difficult for those who don't understand,
and a total mystery to the uninitiated.***

Notes:

Included in this document:

- Nine Basic Rules of Test Creation

- Types of Questions

- ✓ Multiple choice
- ✓ True / false
- ✓ Matching
- ✓ Open-ended

- Types of Tests

- ✓ Performance Based
- ✓ Criterion Referenced

- Closing Thoughts

Notes:

Basic Rules of Test Creation

We'll begin with a discussion of miscellaneous "rules" for test creation.



Do Not Trick Them

The #1 rule in evaluating and testing is DO NOT TRICK THEM. If you have not taught it in the class, it should not be in the test. Forget "You should have read it or know it already." If people are supposed to have pre-requisite knowledge then test them for that, first, before you let them attend the class.

Likewise, you should state your quiz question in the same manner that you stated it/taught it in the class. For example: if you teach the three characteristics of steel, don't ask in your quiz question: *Which one of these is NOT a characteristic of steel* – it's just tricky and mean. Also, for some reason it's hard for most people to have success with "null" answers.



Consider: Why do you want reinforce the *wrong* answer?



Stick to the Facts

The #2 rule is: stick to the facts. Do not include trivial information – the only intention of which is to confuse the test taker.

For instance:

Bob and Ed left their office on "K" Street in Washington DC at 4:45 pm to travel to BWI airport for a 9:00 pm flight – how far is the airport from their office? The times given have nothing to do with the answer you are seeking – in fact, we don't even have to know that Bob and Ed are leaving, do we?

Notes:

3

How Important is the Item?

The more important it is that an item be done correctly – the more test questions should be created for it. You'll want to know – backwards and forwards – whether the person knows his stuff.

4

New Information

Do NOT provide new information in a test question. The test taker is in a search and retrieve mode – he can't be asked to deal with new information.

Example:

Of the 14 types of steels, anodized steel is characterized by what forging process?

If you didn't teach that there ARE 14 types of steel – why introduce it now?



5

Use Key Words

Use the key words: who, what, where, when, how or why to begin questions.

- "Who" triggers the respondent to look for a person or position
- "What" triggers the test taker to look for a thing or a process
- "Where" triggers them to look for a place or location
- "When" will trigger them to look for a date or a period of time
- "Why" will signal them to look for reasoning

Example:

***Who** does Beowulf kill in Beowulf, by **what** method, and **why** does he kill this person?*

Notes:

6

Subjective Responses

Never, never, never ask a question that requires a subjective response.

For instance:

How important is it that the G69 reports are filed daily?

- Very important*
- Important*
- Unimportant*
- Not at all important*



What would be the purpose of asking such a question? It can be assumed that it *is* important; otherwise you wouldn't have been teaching it in the class, right?

Instead, phrase the question as:

In order to avoid getting buried in a mountain of paperwork, it is important to complete and file the G69 reports on a _____ basis.

- Daily*
- Weekly*
- Bi-weekly*
- Monthly*

Or simply,

It is important to complete and file the G69 reports on a _____ basis.

There is no need to provide the rationale for filing the reports on the test.

Notes:

7

Use of Scales

If you are providing a “scale” of choices, make sure the scales start and end at distinct points. For example: If the correct answer is every 3 hours – which response does the test taker check off?

Right	Wrong
<p>A baby should be fed every:</p> <ul style="list-style-type: none"> <input type="checkbox"/> 2-3 hours <input type="checkbox"/> 4-5 hours <input type="checkbox"/> 6-7 hours 	<p>A baby should be fed every:</p> <ul style="list-style-type: none"> <input type="checkbox"/> 2-3 hours <input type="checkbox"/> 3-4 hours <input type="checkbox"/> 4-5 hours <p><i>There are technically two correct answers here.</i></p>

8

Giving Instructions

Instructions are critical. You must do everything you can to make sure test takers know **WHAT** to do, and **WHEN** and **HOW** to do it.

- If there is a time requirement, state it (you must finish this section in 30 minutes).
- If a tool or resource is allowed, state it (you may use a calculator for questions 11 – 20); the opposite is true as well – you may NOT use a calculator to complete this section.
- For each item in column A there is **only one** correct answer in column B.
- If you do not know an answer, it is better to guess than to leave it blank.

It’s also quite helpful to read the instructions out loud at the start of the test – even when they are clearly written on the test. It ensures everyone hears, sees, and interprets the directions the same way and allows you to ask for questions before anyone begins.

Notes:

9

Scoring or Weighting

It is my personal preference that all test questions have the same value. It makes it a LOT easier to grade the test in the end. However, if you feel that certain areas of information that you are testing for are more critical to on-the-job success than others, then you may choose to weight those items more heavily.



Notes:

Types of Questions

Let's look at the different forms of evaluation items and go over some hints and "rules" for each.



Multiple Choice

Typically, when someone is not skilled at writing multiple choice questions, the correct one will always be the longest answer – try to be aware of this and keep your answers to the same length.

It's MUCH easier to provide 3 alternatives than 4; for some reason the 4th one is a killer to come up with.

- **Logic** : Each choice must be logically consistent.

Right	Wrong
Check off all answers that are primary colors:	
<input type="checkbox"/> Red <input type="checkbox"/> Blue <input type="checkbox"/> White <input type="checkbox"/> Orange	<input type="checkbox"/> Red <input type="checkbox"/> Blue <input type="checkbox"/> White <input type="checkbox"/> Pony

- **Difficulty**

Your list of possible choices don't have to be terribly difficult – remember you are testing people who have new knowledge and limited skill/experience – don't come up with answers that would stump *you*, come up with answers that would stump the unskilled person.

By the same token, the right answer shouldn't jump out from the list. You'll want all choices to be somewhat feasible so that the trainee has to put some critical thought in to their answer.

Notes:

■ **Don't Tip Your Hand**

If your answer begins with a vowel, the preceding question should end with "a(n)"; if it ends with *an* – which is grammatically correct – the question itself automatically triggers the trainee to choose the answer that begins with the vowel.

Example: The Interstate is also known as an:

- Expressway
- Highway
- Route

Do you see how ending the question with *an* forces you to choose Expressway?

■ **Parallel Structure**

All your choices should begin with similar phrasing or sentence structure. For Example:

Right	Wrong
The time to cross the street is: A. When the light turns green B. After you have looked both ways C. If the road is clear	The best time for fly fishing is: A. Morning B. Afternoon C. April to May, after the ice pack has melted

■ **All of the Above**

If you're going to provide "all of the above" (or "none of the above") as one of your choices, you should include it in **ALL** of your choices. If you have 15 multiple choice questions and only 1 or 2 have the option of "none of the above," it's a clear signal that the right answer is none of the above.

Notes:

✓ True / False

In true false questions, make sure your options are entirely true or entirely false – do not include answers that might be true under certain conditions.

We tend to write more true statements than false ones, because it is easier; so be aware of that and try to give an equal number of each.

Avoid what are known as “specific determiners” which signal the correct answer – most obvious are ALWAYS and NEVER. Rarely is anything always true or never true. The use of either of these words is usually clue to the test taker to choose the answer that doesn’t include the specific determiner.

Example: “i” *always* comes before “e”



✓ Matching

Matching questions are somewhat easier to create because you only have to come up with one right answer for the “B” column as opposed to the ‘various’ wrong answers you have to come up with in multiple choice questions.

Never provide more than 15 matching questions; if you DO want to provide more, then break them up into chunks of 15 (perhaps that all deal with the same concept). More than 15 really bogs down the process for the test taker. Pretty soon all the letters and numbers are just swimming in front of their eyes; you’ll find that they accidentally reuse the same letters or don’t use one at all. Fifteen is a manageable number of options.

You *may* choose to provide more “B” column answers than “A” column questions. This makes the matching process a bit more difficult for the test taker. Do not provide more than three extra “B” column options, it becomes too confusing.

Notes:

Alternatively, you may have *fewer options* in the “B” column and more questions in the “A” column. In this case you would instruct the test-taker to use the “B” column responses as many times as appropriate.

■ **Example:**

In column “A” identify what type of store layout each named store utilizes. You may use the answers in column “B” more than once.

A	B
1. ____ Home Depot	A. Grid
2. ____ CVS	B. Racetrack
3. ____ JC Penney	C. Boutique
4. ____ Sears	
5. ____ Blockbuster	
6. ____ Shop Rite / Publix	

■ **Provide Logical Wrong Answers**

If you’re going to create a one-of-a-kind question, be *sure* to provide a logical alternative, even if it doesn’t fit any other possibility.

For instance,
A compass always points _____,
 should have the options of at least North and South in the “B” column. If only North is supplied, and you only have one question in the “A” column that has to do with direction, the correct answer is obvious to the test taker.



Notes:

✓ **Open ended questions**

Open ended questions can also be thought of as essay questions or fill-in the blank.

Open ended questions require the trainee to complete the answer using their own words. This type of question allows you to decide if they truly understand the question/situation. It is immensely hard to grade these types of questions; first - it is extremely time consuming, and second - you'll find it quite subjective; even questions that you believe have a straightforward answer will inevitably have some unique responses. Unless there is no other way to determine if the trainee knows the right answers, my last choice would be to administer an open-ended question.



Notes:

Types of Tests

Performance Based Tests

Performance based tests can be done in a classroom environment *or* you may create a test that will go back to the job site and be administered by someone there, who is not a trainer. One of the first items to be defined in this type of test is: Who should be the observer? If the observation is to occur back on the job, and the observer is the person's boss, will the trainee be intimidated? If the observer is a skilled co-worker, will the observer be lenient because it's a buddy? Here are some other considerations when creating observation sheets:



Include instructions for the observer – what should be the setting? Should the trainee be told in advance they will be observed? How long should the observation last? Should the observer give comments/feedback during the performance or wait until the end? Should the observer give comments/feedback AT ALL? Should the trainee be telling the observer what he is doing or should he just do it (e.g. *Next, I will check the depth of...*)?

Your observation sheet will more than likely be a check-off list with yes/no possibilities. The hardest part (for you, the test creator) is breaking down the activity into all of its components. You will have to decide how detailed you want the check-off item to be – for example, if you are testing whether or not someone can change the oil in a car, is it enough to say “locate the oil cap” or should you start with “open the hood”? The deciding factor will be how **important** it is to have the interim steps done correctly or in a specific order. If my ultimate goal is to make a salad – it doesn't really matter which order I cut the vegetables in, so long as they all get into the salad. Likewise I might chop the lettuce with a knife or hand-tear it – there is no “correct” method as it all ends up making a salad in the end. But if I am changing a car's oil it **IS** important to remove the old oil before putting the new in. If I am to relieve the pressure on a gasket **ONLY** if it reaches a certain point, it is **important** that I correctly read the gauge.

Notes:

Criterion Referenced Tests

Criterion referenced tests measure the test-taker against a set of standards. The standards may be level of achievement, speed, quality, a combination of all three, or one of many other types of criterion as well. When scoring a criterion referenced test, the test taker typically either meets, or does not meet, the standard. Other test takers have no bearing on this type of test. For instance, let's say a criterion referenced test has 100 questions and the standard to meet or exceed is 85 correct answers. It does



not matter if I get 90 correct answers and you get 85 – we both have passed the criterion referenced test. A fellow test taker who got 75 correct questions did *not* meet the criterion and therefore failed the test. To pass the Army's PT test a recruit must run 2 miles in 13 minutes. Whether one recruit does it in 10 minutes and another does it in 12 is irrelevant – they have both passed. How one performs on a criterion referenced test predicts how they will perform on the job as well.

These are the only two types of testing that should be done within the workplace. The third typical type of test, the norm-referenced test, compares one test taker to another and unless administered by a professional, could open up your company to all sorts of legal issues. The main use of a norm referenced test is to “classify” people.

Notes:

Combining Types of Tests

It is entirely possible and legitimate to combine performance based and criterion based tests. Pilots must pass a written exam as well as demonstrate that they can take off, land, navigate, and work the controls of an airplane. It's OK to flunk someone who does well on one part of the exam, but poorly on the other. Would you want a heart surgeon who knew his anatomy and physiology inside and out but who couldn't stitch a wound closed? Likewise, there are many people who can **perform** a job but don't really understand the rationale or the process behind why they do the job the way they do it.



Notes:

Closing Thoughts

Ordering the questions

It is really up to you to decide if you want to group your evaluation questions by topic or to mix them up. Personally, I like to mix things up because, back on the job, the work people encounter usually doesn't show up in any kind of logical sequence – so I like for them to have to think “out of sequence” on the exam as well.

The benefit to keeping your questions grouped is that you'll easily spot if a trainee just “didn't get” a particular topic. For instance, if on a cooking class exam, all the sauté questions are grouped together and the trainee gets all or almost all of them wrong – it's apparent he needs retraining in this topic. If your six sauté questions are interspersed throughout the exam however, it might be harder for **you** to spot a problem.

Use of Resources

By resources I mean calculators, policy and procedure manuals, job aids, the training materials themselves, etc. – the rule of thumb here is, if they are allowed to use it on the job, they should be allowed to use it in the evaluation as well. Why make them memorize something they will never need to commit to memory (other than to pass your test)?

Another consideration is this: is your ultimate goal to determine if they got the correct answer or is it how they arrived at the answer? If it's the former, it should be OK to use resources.



Testing the Test

Just as you would have someone else proofread your training materials, have someone take the test as well. They don't necessarily have to have gone through the course or be familiar with it to be able to find the answers – thus ensuring you've taught what you're testing. Remember: if you haven't taught it, it shouldn't be on the test.

Notes:

Notes:

Basic Rules of Test Creation – Job Aid

DO

Stick to the Facts

Create more test questions for items that are crucial

Use key words

Have distinct beginning and end points of scales

Give clear instructions both verbally and in writing

DO NOT

Trick them

Include new Information in a test question

Ask questions that require a subjective response

Ask questions that could be true under certain circumstances

Notes:
