Why Use Performance-Based Testing?

Why Test at All?

James R. Parry, M.Ed. CPT
Owner/Chief Executive Manager
Compass Consultants, LLC
References


Acknowledgments

• Photos, some illustrations and background illustrations used fall within the creative commons license
• Stick figures courtesy of E-Learning Heroes
Why Do We Test?
Testing, or assessment, allows us to validate what an individual knows or doesn’t know.
How Do We Test?
Types of Tests

• Knowledge-based

• Performance-based

• Performance
Knowledge-based Tests

• Assess useful knowledge about required performances
• Must be written to ensure validity and reliability
• Scoring is pretty straight forward - right or wrong - based on a criteria
Performance-based Tests

• A hybrid form that attempts to measure job skills using a combination of knowledge and performance in the same test

• These types of tests are very difficult to write and most test writers don't want to take the time required to develop true performance-based items
Performance Test Defined

"A test or assessment in which the response modalities required to demonstrate competence are identical to or closely approximate the response modalities required to apply that competence in its target context."

The Performance Testing Council (www.performancetest.org)
Performance Tests

• Seek to provide an objective rating of either a behavior or a product
• Assess the demonstration of required skills
• Developing a true performance test is not difficult but establishing the scoring and reliability of the raters is a challenge
Test Forms

CRT vs. NRT
Criterion-Referenced Test (CRT)

• Compares people against a standard
• Test is composed of items based on specific objectives or competency statements
• Defines the performance of each test-taker without regard to performance of others
• Defines success as being able to perform specific tasks or competencies
  • Tests mastery of a concept
• Scores typically high and clustered together
Norm-Referenced Test (NRT)

- Compares people in relation to the test performance of one another
- Composed of items that will separate the scores of test-takers from one another
- Used to rank-order to select top performers
- Does not verify what a test taker can actually ‘do’
- Test scores should be widespread
  - Increases confidence for comparison
Points to Remember

• Criterion-Referenced test is used to test “mastery” of a concept
  • Scores should be high if the training material and tests are doing their job

• Norm-Referenced test is used to rank order (Rack-and-Stack) people
  • Scores should be spread evenly along the spectrum
Food For Thought

• Did the trainee just spend 12 weeks in a school and learn what they were supposed to learn and are they able to perform satisfactory on the job?

• Can we diagnose trainees or even instructors’ deficiencies?

• Can we say with confidence that a trainee has mastered the required objectives or job tasks?
When Do We Test?
Formative Assessments

• Given at intervals during the course to test the knowledge and skills being formed within the students
• Tests given during courses and schools are based on the objectives
• We are testing the knowledge and abilities of individuals to meet or perform these objectives
• Helps student recognize strengths and weaknesses
• Helps faculty recognize where students are weak so problems may be addressed
Summative Assessments

• Goal is to evaluate student learning at end of instruction by comparing it to some standard
• Often high stakes
• Information obtained can be used to modify instruction in subsequent offerings of course
What do the results of written assessments tell us?
What They Know
If the objectives are written to only “know” something then all is well!
Useful Objectives

• Three characteristics
  • Performance
    • Describes what the learner is expected to be able to DO
  • Condition(s)
    • Describes the conditions under which the performance is expected to occur
  • Criterion
    • Describes the level of competence that must be reached
“Objectives increase the accuracy with which cognitive processes in particular can be assessed. A well-written objective is the blueprint for the creation of test items that will assess the specific competency described by the objective. In this way, objectives make it much easier for test writers to create so-called parallel test items, that is, different test items that assess the same objective.”

Shrock & Coscarelli, (2007, pg. 106)
Can a performance objective be fully tested with a knowledge-based assessment?
Knowledge Tests

• Usually formative
  • Measure retention of key concepts or knowledge during training
  • Used as building blocks to performance

• May be used to predict readiness to demonstrate performance
Basic Knowledge Only Test Item Example

1. In which octave is the key used to sound “middle C” located on a standard 88-key piano keyboard?
   A. In the 3rd octave
   B. In the 4th octave
   C. In the 5th octave
   D. In the 6th octave
Limitations of Traditional Knowledge Tests

• Include the common 4-alternative multiple choice, multiple response (several alternatives must be selected), matching, etc.
  • Cannot test actual performance

• Most times, these types of test items are written entirely at the knowledge level where the respondent regurgitates chapter and verse from learning material

• We don’t know if the test taker can actually ‘do’ anything
Performance Tests

• Diagnostic/evaluative/summative
• Used to evaluate actual performance
  • Must be designed to test in actual conditions encountered on the job
  • Must use actual equipment
  • Little or no verbal or written testing
Performance test item example:

- Dissect a preserved frog and remove the stomach, heart, lungs and reproductive organs following the checklist provided without error.
Performance Checklists

• Used to ensure all required steps are followed
• Provide standard for all observers/graders
• Ensures each participant is graded to same standard
• Questionmark® Assessment Management Systems (Perception & On-Demand) offer “Observational Assessment” capability
Advantages of Full Performance Testing

• Allows us to test what an individual can actually do either in a classroom environment or on the job and not just what they know.

• Helps us measure an individual’s competence by having them perform a task or set of tasks in an actual work environment while being observed and graded to an acceptable level of performance by a competent, qualified trainer or observer.
Limitations of Performance Testing

- Cost of purchasing sufficient equipment for each trainee is often cost prohibitive
- Time consuming
- Little time for retesting
- Full performance testing of large class is sometimes difficult
  - Student/observer ratio
- Group performance testing does not ensure all can perform
The Hybrid Approach – Performance-Based

• By combining elements of both traditional testing methods with performance testing it is possible to test at a higher level of fidelity if actual equipment or real-life environments are not feasible.

• Hybrid approach ensures the trainee has attained the knowledge required to be able to perform a task before actually performing it, then incorporates traditional testing at a higher cognitive level with simulated performance.
  • Then test only those actual performances that cannot be simulated by higher level traditional testing.
Performance-Based Testing

- Combination of application of knowledge and performance testing
- Allows us to test what and individual can apply either in a classroom/lab environment or on the job and not just what they know
  - Measures competence
  - Allows observation by qualified instructors
- Trainee is told to, “Apply what you have learned by showing what you can do.”
Performance-Based Example

Drag and drop the appropriate schematic symbols to build a DC circuit that will provide continuous power to a lamp unless the power is interrupted by a pushbutton switch. (Note: Some symbols may not be used. Drag unused symbols to the box below the schematic diagram.)
Limitations of Performance-Based Testing

• May not be at true fidelity encountered on the job
  • May not fully test actual performance

• Sometimes erroneously used to test true (full) performance objectives
  • Written at on the job level but not obtainable in the classroom

• Example:
  • Performance objective: CONDUCT a full orchestra in an amphitheater with the performance of the score from *Priscilla, Queen of the Desert* without error
  • Erroneous Performance-Based test: Compare and contrast, in 100 words or less, the woodwind vs. string portions of the score from Priscilla, Queen of the Desert

Copyright © 2017 by Compass Consultants, LLC
“Skills and knowledge are required in any job. If they are missing, job performance is impaired and training may be required.”

Rummler & Brache (1995, p. 72)
When, What, How and How Much?

• How do we decide when and how much of each method to use?
• How about the difficulty and level of complexity of test items?
• When should we stop traditional testing methods and move on to performance-based?
How can I put this all together?
Determine the Mix

• Solving the “balancing act” begins with integrating key elements into a well written test plan
The Test Plan

• This is the blueprint that will guide when, at what cognitive level and difficulty, and how testing will be accomplished

• Development of a test plan will also reveal any deficiencies in the construct of the objectives

• If you can’t decide on a way to test or develop several test items for an objective – it may not be a solid objective – it could just be a step to a larger objective
Objectives – The Most Important Key Element

- Learning objectives must be written with job performances that are obtainable in a training or laboratory environment.
- If testing on the ‘real deal’ is not possible, the trainee should be tested at the highest level of simulation possible given the constraints imposed by the training environment.
- Once the objectives are written to support the level of obtainable performance, a plan for testing must be developed.
Course Designers May Not be Test Designers

- Sometimes objectives, as written, do not work well in testing world
- Important to communicate!
- Develop test items immediately after objectives are developed
- If test designer can’t come up with 3 – 5 measurable test items per enabling objective (EO) there may be a problem
  - May be written at the step level
Poorly Written Objectives

• Terminal Performance Objective
  • 1.1 – Given all required components, FABRICATE a hydraulic hose without error

• Enabling Objectives
  • 1.1.1 – MEASURE 24 inches on a 3/16 inch hose
  • 1.1.2 – APPLY a piece of tape at the 24 inch mark
  • 1.1.3 – And so on...

• This terminal performance objective is more in line to be an enabling objective of a larger performance and the enabling objectives shown are steps of FABRICATE
Overt vs. Covert Performance

If you can’t see it how can you test it?
Overt vs. Covert Performance

• There are two types of performance
  • Overt (visible) performance
  • Covert (invisible) performance

• Both are performances that are measurable
  • Covert may be difficult to measure
Overt Objective

• Given an electrical circuit with a power source, various loads and a clamp-on ammeter, MEASURE the amperage draw of the circuit within ± 0.1 amps

• The performance – MEASURE – is something that can be observed, a covert action; the conditions are clearly given – a power source, various loads and a clamp-on ammeter; and the criterion is that the measurement obtained must be within ± 0.1 amps
Testing Overt Performance

• Pretty straight forward
• Use formative testing as knowledge check
• Increase complexity and difficulty until performance is demonstrated
Covert Objective

• Given an electrical circuit schematic with a power source and various loads indicated, CALCULATE the amperage draw of the circuit within ± 0.1 amps

• The performance is still there – CALCULATE – but the act of calculating may not be something that is easily observed if the performer completes the calculations without any physical device such as a calculator or pencil and paper
Testing Covert Performance Objectives

• How can we be sure the performer can actually calculate the amperage?
• What type of performance test can we use?
• How about a test written at various cognitive and difficulty levels that will SHOW us the trainee can DO the required performance
At what level should tests be written?
Bloom’s Taxonomy

High
- Evaluation
- Synthesis

Moderate
- Analysis
- Application

Low
- Comprehension
- Knowledge
Low Complexity Testing

The Knowledge and Comprehension levels of the Taxonomy lend themselves well to traditional knowledge testing methods – the “Do you know it?” phase using low complexity test items.
Moderate Complexity Testing

The Application and Analysis levels can be tested with items that use application of knowledge – the “Can you apply your knowledge?” phase using test items written at a moderate level of complexity that make the trainee think and use their knowledge in new situations.
High Complexity Testing

Finally we move up to the Synthesis and Evaluation levels which can become the, “Show me your stuff!” part of testing using high complexity test items and actual hands on performance.
How can all of this be balanced?
Can You Apply Your Knowledge?

Do You Know It?
- Application
- Comprehension
- Knowledge

Show Me Your Stuff
- Evaluation
- Synthesis
- Analysis

Hybrid of Both

Traditional Testing Methods

Performance-Based and Performance Testing

Copyright © 2017 by Compass Consultants, LLC
Summary of Performance

• We need knowledge to perform
  • Knowledge-based testing

• We need to be able to apply knowledge to make sure we can perform
  • Performance-based testing

• We need to be able to show that we can perform correctly
  • Performance testing
So...

Why do we need to test?
Questions?
Thank You!!!

Website: [www.gocompassconsultants.com](http://www.gocompassconsultants.com)

E-mail: [james.parry@gocompassconsultants.com](mailto:james.parry@gocompassconsultants.com)