Performance Item Beta-Testing Process

Wallace Judd, PhD

Performance Item Beta-Testing Process

Wallace Judd, PhD

Performance Item Beta-Testing Process

Description

Beta testing is the formal process of initial evaluation of items written for the test being developed.

Beta test is the initial, informal evaluation of an item. It can be done with several trial subjects who would presumably know how to do the task correctly. The goal is to see whether the item works, and what elements may not have been included in the initial item plan.

Purpose

The purpose of the pilot test is to evaluate the quality of the questions and the appropriate time limits for completion of the test. The results of the pilot test will be used to (1) select the final scored items in the test, (2) determine a test passing score, and (3) set the appropriate time limit for the test.

These three purposes make it clear just how critical the beta-testing process is to exam development.

Subjects

Selecting appropriate subjects is of critical importance in the beta-testing process.

Because the beta results will be used to establish a cut score for the exam, it is critical that your subjects be representative of the candidates you wish to evaluate.

If your subjects are too expert, then the cut score will be above the level of competence necessary to perform expected tasks.

If your subjects are unprepared, naïve, or inexperienced, the cut score you establish will likely be lower than needed to show appropriate competence.

To obtain the right distribution of candidate talent, Table 1 provides helpful guidance.

Where to Get Candidates

Low levels   Applicants with little or no hands-on experience; ones who have applied for training but have not completed it.

Target distribution of candidate skills for a beta test.

Candidate Distribution Targets		
Fraction	**Skill Level**	**Skill Description**
⅙	Low skills	below the preparedness expected of competent candidates
⅔	Medium skills	candidates with appropriate skill level for the job
⅙	High skills	expertise above that expected of competent practitioners

Medium level Candidates who have completed the appropriate training and have some experience on the job, but who haven't practiced more than six months to a year on the job.

High level Candidates who mentor others in the skills being evaluated. People who have extensive on-the-job experience in several roles. Leaders in the industry.

Timing

One purpose of the beta test is to establish an appropriate time limit for the test. This is a more delicate process for performance tests than for multiple-choice tests, unless the multiple-choice tests involve substantial computations.

The beta test should be administered with no time limits. If it is longer than 3 hours, then candidates should be given a break and resume testing for a final segment no longer than 3 hours. If the test requires more than 6 hours, then it should be administered over two or more days of testing.

After a cut score has been established, and the items for the test finalized, then an appropriate time limit can be set. One helpful formula for establishing a test time limit is:

$$\text{Time Limit} = (MT + 2 \times Sdt)$$

where

MT is the mean time used by passing candidates, and

SdT is the standard deviation of the time used by passing candidates.

This ensures that approximately 95% of the candidates will have sufficient time to complete the test.

The time limit above assumes that approximately 10% of items on the final test will be trial items.

Item Composition

Items incorporated in the test must reflect the domain weightings specified in the blueprint, which is defined in the scheme.

Each of the test domains should have a minimum of 10 percent more items included than are provided for in the blueprint. These additional items are trial items that may be used as scored items if some of the

items initially intended to be scored turn out to have inappropriate parameters.

Analysis

Conduct psychometric analysis to evaluate the performance of items included in the beta test.

Minimum item parameter calculations are the p-value, point-biserial correlation with overall score (excluding the item), item time, p-info, and item efficiency. (See psychometrics essays in this series.)

Once the parameters are evaluated, scored items can be selected for administration in the final test form or forms, either as scored or trial items, and a cut score established.

Outcome

The completed beta should have the following outcomes:
- ☐ A cut score for the final test
- ☐ Determination of scored and trial items to be included on the test
- ☐ Test administration time limit
- ☐ Candidate pass/fail scores on the final exam calculated and reported