Gating Items: Unique to Performance Tests

Wallace Judd, PhD

07

Gating Items: Unique to Performance Tests

Wallace Judd, PhD

07

Gating Items: Unique to Performance Tests

Overview

A gating item is an item that must be passed in order to pass a test. It may be presented at any time during the course of an exam. Whenever it is presented, it is a prerequisite to passing the test.

Gating items are a natural, often inevitable result of the conditions of performance.

I was in Montana in the summer of 2007 and saw numerous ospreys fishing. Osprey nests are high in trees or perched high on telephone poles. Clearly, fledgling osprey are presented with a gating item: the first time they launch themselves from the nest they must fly. Otherwise they plummet to earth and kill or injure themselves. Their first flight is a gating item in the test of life.

Examples

Gating items really exist in certification exams as well. The extensive examples presented below are intended to present evidence that gating items not only exist; they are natural and intrinsic to the content being tested.

In writing a Microsoft Word performance test that was administered in April, 2007, I inadvertently included two gating items. Only subsequent analysis revealed that they were included in the test.

Candidates were asked to open a file, edit the file according to a template document, then save the file with a new name so it wouldn't overwrite the source document. The scored content of the item was the edits the examinees made to the file. What I didn't realize was that opening the file and saving it under a new name were gating items. If examinees can't save the file, it doesn't matter how well they edit the file. In fact, if they can't save files correctly, it doesn't much matter how well they can edit in general, since those edits won't be saved. Clearly, if someone can't save a file in Word, no matter what else their skills are, they can justifiably be failed.

Similar items exist on more sophisticated exams. In an exam of Linux system administrators, for example, candidates who can't add a new user to the system will fail the exam no matter what other skills they exhibit. Obviously, no matter how well candidates can install printers, load balance, or tune the system, all the benefits their skills deliver are useless if new users can't access the system.

Figure 1. A carpenter saws a wooden beam without protective equipment. In a performance test, use of personal protective equipment is likely a gating item. The candidate would fail immediately, notwithstanding any carpentry skills he may possess.

Likewise, on an exam of Oracle system administrators, backing up the transactional database is a gating item. During the exam, a catastrophic failure of the operating system is induced. If the candidate hasn't backed up the system with each transaction, the system crash becomes an unrecoverable event. This is simply unacceptable for a competent database administrator. Again, no matter what other skills are exhibited, the candidate justifiably fails and cannot continue the exam, since any other skills are overwhelmed by the lack of a system backup.

In a private pilot's practical flight test administered by the Federal Aviation Agency, inability to land the plane in three attempted landings results in failure on the exam. No matter how well the pilot may plot a course, conduct the takeoff, fly the route, or plan the approach, if the pilot cannot land the plane in a reasonable number of tries, the examiner takes over, lands the plane, and fails the candidate. And that's clearly as it should be. What reasonable person would allow a novice pilot to take off if that pilot did not demonstrate the ability to land the plane?

A final example may clarify the issue. Foreign-trained veterinarians are given a practical exam by the American Board of Veterinary Medicine before they are certified to practice in the United States. The practicum includes seven stations at which examinees treat live animals exhibiting a variety of symptoms and requiring a variety of treatments. At one of the stations, candidates are asked to spay a cat or dog. If the candidate puts the animal's life in jeopardy, the attending veterinarian takes over and tries to save the animal, and the candidate discontinues practice and has failed the practicum. This is clearly a gating item.

The previous documentation includes such an extensive number of

Figure 2. A worker climbs a telecommunication tower without a safety harness and safety equipment. In a performance test, this action is likely to result in an immediate fail. *See facing page.*

Figure 3. A system administrator candidate who neglected to back up a transactional database fails a performance test because they cannot recover from a catastrophic event induced as part of the test. *Right.*
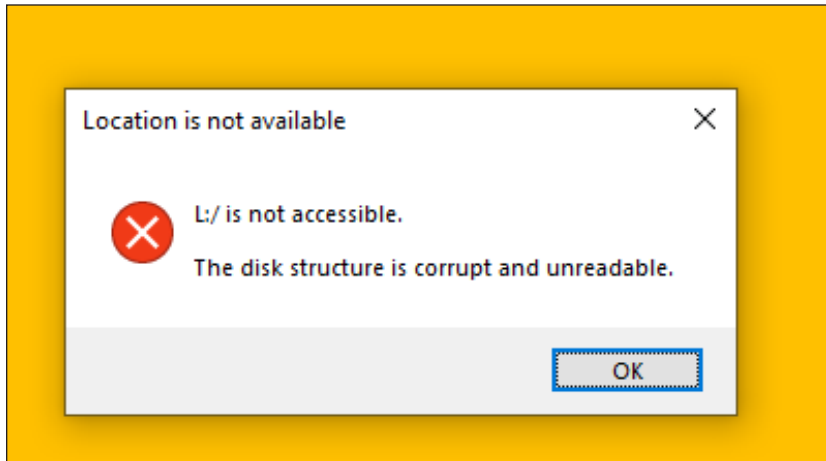
examples of gating items to clarify for the reader that such items do actually exist and represent appropriate, indeed natural, scoring.

Why Gating Items Have Not Been Recognized

It is important to note that gating items would typically not arise in multiple-choice exams for several reasons.

(1) In a multiple-choice exam, the examinee can continue from any incorrect item. In a performance test, an incorrect performance could reasonably terminate the exam. For example, if the candidate can't open a file in Word, how can they be asked to perform subsequent edits on that file? In a concurrent test of the operating system, the examinee could possibly delete the operating system and prematurely terminate the test.

(2) In a multiple-choice exam, complex events are usually represented by multiple items, any one of which doesn't appear to be catastrophic. In a performance exam, complex responses frequently have a single, scoreable result. If a veterinarian doesn't perform an acceptable sequence of complex actions, the dog may die.

(3) In many multiple-choice exams, the correct response may be obvious. When asked, any database administrator would say that the database should be backed up not weekly, daily, or hourly, but after every single transaction. Yet a candidate may fail to exhibit that behavior when actually performing a system configuration. The difference between what a candidate articulates and what the candidate does becomes immediately apparent during a performance test.

Theoretical Problems Posed by Gating Items

Gating items present significant difficulties for Classical Test Theory.

How is one to compute a meaningful point-biserial coefficient? The result from the classical formula (Guilford, 1965, p. 323)

$$r_{pbi} = \frac{M_p - M_t}{\sigma_t} \sqrt{pq} \qquad \text{(Eq. 1)}$$

is zero when the test is terminated because of failure on an item, since

in that case $M_p = M_t$. In the above equation,

M_p = Mean score of those getting the item correct
M_t = Mean of all test scores
σ_t = Standard deviation of test scores
p = Proportion of people getting the item correct
q = Proportion of people getting the item incorrect

Under many testing guidelines, an item with a reported point-biserial coefficient of zero would be dropped from the exam form. Yet, as the numerous examples cited above illustrate, gating items form a critical part of many performance exams.

How would one evaluate equivalent forms? If a gating item is so critical, is it reasonable to create an equivalent form of the item?

How is one to account for the test scores of examinees who are allowed to continue after failing a gating item? Are these legitimate scores that should be included in the mean score? Or are they only valid as pass/fail scores? How do you report to a candidate that a test score above the cutpoint resulted in failure on the exam?

What if an examinee has taken a portion of the exam prior to encountering and failing a gating item? Is the examinee's test score the score for the portion of the exam the candidate was allowed to take?

The Angoff cutpoint determination procedure has no contingency for dealing with gating items. Angoff requires that minimum competency levels for all items be averaged, and the average multiplied by the number of items on the test to determine the passing score. This process does not work with gating items.

What is the proper procedure for computing Cronbach's *alpha* when an exam includes a gating item? If the exam was terminated on failure of the item, clearly the correlation of the gating item with all subsequent items is undefined. If the exam is continued after the gating item is failed, computation of *alpha* is possible but meaningless, since passing or failing subsequent items is meaningless.

These issues present serious unresolved issues for Classical Test Theory and Item Response Theory to deal with.

Fortunately, in actual practice, gating items are infrequent, typically comprising a small proportion of the items on a test form. Moreover, they are so fundamental to the practice being evaluated that they typically have a low observed frequency of failure. Nonetheless, they are a legitimate component of performance testing, and need to be incorporated within the theoretical framework of standard setting, exam evaluation, and item evaluation.

Implications for Practice

Despite the unresolved theoretical issues discussed above, gating items will continue to appear in performance exams. For those developing performance exams, a few suggestions follow.

(1) Recognize gating items that are intrinsic to the content of the exam. Do not avoid them or, conversely, try to find them if they are not an inevitable part of the domain being tested.

(2) Do not eliminate gating items because their *p*-values are high. A *p*-value of 0.95 to 0.98 is quite normal for a gating item.

(3) Include the gating item as an explicit part of the scoring rubric. One could conduct an Angoff evaluation of exam items, excluding gating items, then state a scoring rule in the form:

A passing score consists of passing all gating items and achieving a 78% score on the remaining items.

(4) Be sure that the directions for the gating item are absolutely clear. These items must be unambiguous because they have the most severe consequences of any items on the exam.

Citations for Gating Items

It is difficult to find citations for gating items since they have not been documented by academic studies of testing. Cizek and Bunch (2007) recognize nine different methods of standard setting for exams; gating items are not mentioned in any one of the methods. The comprehensive treatment of Item Response Theory by Baker and Kim (2004) neglects to mention gating items. The only mention of gating functions appears to be in multistage testing where the results of one tier may gate access to another stage of testing.

Terminology

Discussions with other psychometricians and practitioners have turned up several terms for these items: 'domain-critical items,' 'critical items,' 'mandatory items,' and 'gating items.'

The terms 'domain critical items' or simply 'critical items' do not convey the absolute, unmitigated requirement for passing that the term 'gating items' connotes. *Criticality* is a matter of degree similar to *importance.* As a matter of practice, criticality is often evaluated as a scalar in job task analysis.

The term 'mandatory items' calls to mind the 'mandatory' figures an Olympic skater must perform as part of the suite of routines performed for the judges. Skating the figures is mandatory; scoring perfectly on them is not mandatory. Hence there is little parallel with these item types.

The term 'gating items' seems particularly apt when reviewed in light of a logic gate: it is an AND gate. The examinee must pass this item AND any other set of items required for passing. Hence, I refer to these items as gating items.

Works Cited

Baker, Frank B., and Seock-Ho Kim. *Item Response Theory; Parameter Estimation Techniques.* Second Edition. Marcel Dekker, 2004.

Cizek, Gregory J., and Michael B. Bunch. *Standard Setting, A Guide to Establishing and Evaluating Performance Standards on Tests.* SAGE Publications, 2007.

Guilford, J. P., *Fundamental Statistics in Psychology and Education.* Fourth Edition. McGraw-Hill, 1965.