## CSig 2color smallPerformance standards and the standard setting process

**Castle Worldwide, Inc.**

**What is a performance—or criterion-referenced—standard?**

In contrast to a norm-referenced assessment, a criterion-referenced standard compares a candidate’s performance against an objective performance standard, rather than against the performance of other candidates. This requires a candidate to show specific knowledge, rather than a performance that is compared to other candidates. Upon successful completion of a criterion-referenced test, a candidate demonstrates a link between the purpose of the credential and an established standard of competence. Criterion-referenced examinations rely on a performance standard that is translated into a cut-score, which is the score required to pass an examination.

**How does a credentialing organization establish a performance standard?**

The process begins with the organization’s examination development team carefully selecting a panel of subject matter experts (SMEs) who possess intimate knowledge of the capabilities of the intended candidate pool (e.g., through supervision and/or education of potential candidates). The SMEs review what is known as the “base” or “anchor version” of a test—that is, the standard to which later versions of the test are anchored. From the perspective of a minimally competent test candidate, the SMEs assess the level of difficulty of each item on the examination, or of the test as a whole.

Beginning in the 1950s, a process known as “standard setting” was developed to establish the performance standard of an examination and to set a meaningful passing standard. The passing standard is the minimally acceptable performance required by a candidate.

The SMEs convene for a standard-setting meeting. Different methods exist to conduct standard setting studies, but the most common method is the Angoff family of techniques. For these methods, the SMEs are trained to:

1. reach a common understanding of the purpose of the examination (including the intended test takers);

2. reach a common understanding of the standard of minimally competent performance, which comports with the purpose of the examination; and

3. estimate the difficulty of each item for a minimally competent candidate.

The SMEs in an Angoff-type standard setting meeting perform an iterative process, whereby they typically review the test two or three times and examine the consequence of their decision. The purpose is to refine the accuracy of the identified passing standard. There is no desired percentage of successful candidates, nor is there a fixed percentage of correct answers that influence the standard.

**How often should a credentialing organization establish a performance standard?**

Testing standards do not provide specific recommendations for the frequency of establishing a performance standard. However, statistical processes, such as Item Response Theory, assume that test candidates are similar over time (i.e., sample invariance) and that the content being tested is relatively stable.

**How often do credentialing organizations review the performance standard?**

Some organizations designate a specific timeframe for establishing a new passing standard. However, such an approach may restrict the program’s ability to respond to changes in the community of test candidates or requirements of the program.

An annual review of the relevance of the existing performance standard is recommended, typically following delivery of the annual report. The purpose of the annual review is to assess:

1. whether changes have occurred in the population of test candidates (for example, a significant increase in retake candidates or changes in the candidates’ educational/work experiences);
2. accuracy of the existing equating (for example, based on review of the current and historical passing rates); and
3. item performance data and evidence of changes in the candidates’ capabilities.

The review panel determines whether the existing standard remains relevant, or identifies a timeframe for establishing a new standard.

**Do all test forms undergo this process?**

Not all test forms must undergo a standard setting process to establish a passing score. Credentialing organizations must ensure that all test forms are comparable in difficulty so that candidates who take different test forms on different dates are treated equally. However, each item on a test form differs in its level of difficulty. In result, some test forms are harder (or easier) than others. An equating process can be used to account for differences in test form difficulty.

Most organizations use an equating process that involves the common-item nonequivalent groups design. This design compares the performance of one group of candidates on a test form to another group of candidates on a different test form. The two groups of candidates are assessed for how well they performed on the items in common between the different forms. The assumption is that the groups will perform similarly on the common items, and that any difference in difficulty between the test forms is due to the difficulty of the items that are unique to each form. Adjustments can then be made to account for any difference in the overall performance of the candidates.