

Standard Setting/Process of Determining a Cut Score

The passing score or cut score for many high-stakes certification programs is established through a psychometrically valid standard setting process. Meara et al. (2000) note that “One of the most critical and controversial aspects of the credentialing examination process is standard setting. The process...is critical because the pass-fail decisions can have tremendous consequences for the examinees...Passing scores, for the most part, are judgmental in nature. They do not reflect any universal truth” (Meara, Hambleton, & Sireci, 2000). However, even if cut scores are not universal truths, it is important to document that an acceptable, systematic, and thoughtful process was followed in arriving at the recommended cut score standard (National Commission for Certifying Agencies, 2014). Hambleton indicates that “the defensibility of the resulting standards is considerably increased if the process reflects careful attention to: (a) the selection of panelists; (b) the training of panelists; (c) the sequence of activities in the process; (d) validation of the performance standards; and, (e) careful documentation” (Hambleton, 2001, p. 93).

Standard Setting Methodology

Standard setting methodology can be broadly categorized into two approaches: norm-referenced and criterion-referenced.

A **norm-referenced approach** uses the performance of a sample of examinees to determine the passing score of an exam. The pass mark is usually set as the mean (or the mean minus some number of standard deviations) of the candidate group. Thus, the success of a candidate is dependent upon the performance of others who previously took the exam. Using this approach, the cut score would be set so that a pre-determined percentage of candidates pass and fail. Many college entrance exams and national school exams (e.g., SAT, GRE, etc.) use the norm-referenced approach. The primary advantage of the norm-referenced approach is that they can provide information on how an individual's performance on the exam compares to others in the reference group. One of the disadvantages of this type of approach, however, is an inability to identify which candidates can correctly perform the tasks at a level that would be acceptable for employment. As such, they do not present a direct link between competencies required for successful job performance and passing.

A **criterion-referenced approach** uses exam scores to generate a statement about the behavior that can be expected of a candidate with that score. Most credentialing exams can be considered criterion referenced. In this case, the objective is simply to see whether the candidate has learned the material. The purpose of a criterion-referenced standard is to determine "how much is enough" to be considered capable to practice in the field. The validity of the pass or fail inference depends on whether the standard for passing makes a valid distinction between acceptable and unacceptable performance. The level of performance required for passing should depend on the knowledge and skills necessary for acceptable performance and should not be adjusted to regulate the number of candidates passing the exam. These standards suggest that an absolute expectation for passing should be established based on the exam content, prior to administration. When a criterion-referenced standard is established, the candidates who meet the standards pass the exam, and those candidates who fail to meet the standards fail the exam.

A criterion-referenced approach is well suited for adaptive exams, given the nature of such exams. Adaptive exams are designed to measure candidates' knowledge of content areas, while adapting to their ability levels. These computer-administered exams proceed by selecting subsequent test questions depending upon correctness of previously answered questions. Thus, adaptive exams match the exam difficulty to candidates' ability, minimizing measurement error. As a result, no two candidates have an identical exam. Since criterion-referenced cut scores are also defined in terms of required level of knowledge and skills for passing, this approach is appropriate for adaptive exams.

Angoff Procedure of Standard Setting

The Angoff procedure and its derivatives, such as the Modified Angoff procedure (Angoff, 1971), is one of the most used criterion-referenced approaches to setting pass marks on multiple-choice credentialing examinations. According to the 2008 standard setting survey of credentialing organizations whose certification programs were accredited by the *National Commission for Certifying Agencies (NCCA)*, 75% of organizations used the Modified Angoff procedure. When asked about critical factors that influenced their choice of this standard setting procedure, the organizations mentioned the reliability of the procedure, its acceptability regarding the NCCA or ISO/IEC accreditation standards, and the ease of use.

There is considerable evidence that the Modified Angoff procedure is easy to explain and efficient to implement. The participants can easily comprehend instructions for this standard setting procedure, and it takes them a reasonable amount of time to complete standard setting ratings. Research has shown that the Modified Angoff procedure produces reasonable results that meet testing standards for assessment reliability and validity of assessment scores. This procedure offers the best balance between technical adequacy and practicability.

During the process of the Modified Angoff standard setting, a panel of subject matter experts (SMEs) discusses the **minimum level of competence** that is required for passing the examination and obtaining the credential. The panel participants review the content outline established through the process of *job analysis*, representing the knowledge, skills and abilities demonstrated by the practitioner. The subject matter experts utilize the job tasks to develop and review performance-level descriptors (PLDs) that provide information about the typical knowledge and skills of *minimally competent*, *not-yet competent*, and *highly competent* candidates. After evaluating and analyzing the difficulty of each question, as well as the specific knowledge, skills and abilities that qualified practitioners possess, a raw cut score, or the passing score, is set for that particular exam form through multiple rounds of rating assignments and discussion. This becomes the standard for the program. As new exam forms are created, equating is done to adjust the passing score as needed to account for any differences in form difficulty.

Once a standard/pass mark is established, its relevance must be periodically evaluated. Any time a Job Analysis Study is performed, or the examination content is altered, a new standard setting study must take place to re-establish a passing standard which is aligned to the new content outline.

Scaling and Equating

Scaled Score

A scaled score is a representation of the total number of correct questions a candidate has answered (raw score) that has been converted onto a consistent and standardized scale. Scaled scoring is a certification industry best practice for reporting high stakes exam scores, used to account for the potential differences in difficulty across unique exam forms.

Form Differences and Scaling

For fair and consistent decisions to be made on exam results, scores should be comparable. This means that scores from different forms of a test should indicate the same level of performance regardless of which exam form a candidate has received. This will take into account the potential variability in difficulty between different exam forms. Test developers adhere to strict test specifications when developing multiple exam forms to ensure that they are similar in difficulty. Because of the variability in difficulty of individual questions, the forms may vary in difficulty. For this reason, percent-correct scores do not always represent a fair comparison of different forms. For example, a candidate scoring 50% correct on a very difficult exam form would likely have more knowledge and skills than a candidate scoring 60% on an easy form. For the same reason, raw scores cannot be used, as two candidates with the same raw score on different forms would have demonstrated different levels of performance relative to the difficulty of the individual exam forms. A scaled score provides a standard range for candidates and allows direct and fair comparisons of results from one exam form to another.

Scaled scores DO NOT affect individual candidate pass or fail decisions. The decision is always made by comparing the number of items answered correctly to the number of items required to pass the test form. Candidates who correctly answer enough items to pass the test form will obtain scaled scores between the passing scaled score and the maximum scaled score. Candidates who do not answer enough items to pass will obtain scaled scores from the minimum scaled score to just below the scaled passing score.

Equating

Candidates are assured fairness when form difficulty varies by a statistical process called equating. Equating procedures measure the difficulty of each exam form and adjust the passing score as needed so that the same level of candidate performance is reflected in the passing score regardless of the difficulty of the form. By using equating procedures, an equivalent passing standard for each form is maintained. Candidates who happen to take a slightly more difficult exam form are not penalized. Likewise, candidates who take the slightly easier exam form are not given an unfair advantage.

The table shows an example of scaled scores associated with different raw scores for two different exam forms, Form 1 and Form 2.

Raw Score	Scaled Score	
	Form 1	Form 2
100	305	305
99	304	303
98	303	301
97	302	300
96	301	299
95	300	298

As seen in the table, Form 1 is the more difficult form because it requires a smaller number of correct questions to achieve a passing score of 300. The minimum raw score required to achieve a scaled score of 300 on Form 1 (95) is different from the minimum raw score required for Form 2 (97). The passing score on each of these two forms, however, is reported as the same number; that is, 300.

References

- Angoff, W. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (ed.), *Educational Measurement*. Washington, DC: American Council on Education.
- Hambleton, R. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. In G. Cizek (ed.) *Setting performance standards: Concepts, Methods, and perspectives*, (pp. 89 – 115). NJ: Lawrence Erlbaum Associates.
- Meara, K., Hambleton, R., & Sireci, S (2000). *Standard-setting practices in the credentialing examination field*. NY: Professional Examination Service.
- National Commission for Certifying Agencies (2014). *Standards for the Accreditation of Certification Programs* (pp. 20-21). Retrieved November 14, 2018, from:
https://cdn.ymaws.com/www.napo.net/resource/resmgr/Docs/2016_NCCA_Standards.pdf