

Measurement Research Associates

Criterion Referenced Performance Standard Setting

The level of performance required for passing a credentialing test should depend on the knowledge and skills necessary for acceptable performance in the occupation or profession and should not be adjusted to regulate the number or proportion of persons passing the test. The pass point should be determined by careful analysis and judgment of acceptable performance.

There are many different systematic methods used for selecting passing standards for examinations. Seven methods are presented.

Angoff's Method

The passing score is computed from an estimate of the probability of a borderline candidate answering each item correctly. After a discussion and consensus of the characteristics of a borderline candidate, each judge makes an independent assessment of the probability that a *borderline candidate* will answer the item correctly for each item. The judges' assessments of an item are averaged to determine the probability of a correct response for that item. Then, each probability assigned to an item on the exam form is averaged to obtain the pass point. The benefit of the Angoff is that it has held up in court, is relatively straightforward, and does not require exam data.

Modified Angoff's (Yes/No) Method

After a discussion and consensus of the characteristics of a borderline candidate, judges review each item to answer the question: "would a borderline candidate be able to answer this item correctly?" The items they should answer correctly are assigned a 1 = yes, and the items they should not answer be able to answer correctly are assigned a 0 = no. The pass point is then calculated by averaging the scores. The modified Angoff has been described as much easier than estimating the proportion correct as used in the traditional Angoff.

Nedelsky's Method

The method is based on the concept that the borderline candidate responds to multiple-choice items by first eliminating the responses he/she believes are incorrect and then assessing the plausibility of the remaining responses. Borderline candidates should be able to reject incorrect answer responses. To establish the pass point, judges independently identify the answer option(s) that a borderline candidate would be able to recognize as implausible. The number of remaining options determines the probability that the candidate will answer correctly the item: 1 plausible response = 100%, 2 = 50%, 3 = 33%, 4 = 25%, and 5 = 20% probability of a correct response. The average of the probabilities determines the pass point. This method can only be used with multiple choice items.

Ebel's Method

Ebel's approach is accomplished in two stages and takes into account the item's difficulty and the item's importance. Judges independently classify item difficulty as easy, medium, or difficult, and importance as questionable, acceptable, important, or essential. This forms 12 groups of items. The

characteristics of the borderline candidate are discussed. Judgments regarding the percentage of items a borderline candidate should answer correctly in each of the 12 categories (3 difficulty x 4 importance) is made. In each category, the percentage correct is multiplied by the number of items in that category. Category scores (percentage correct X number of items) are averaged to determine the pass point.

Modified Ebel's Method

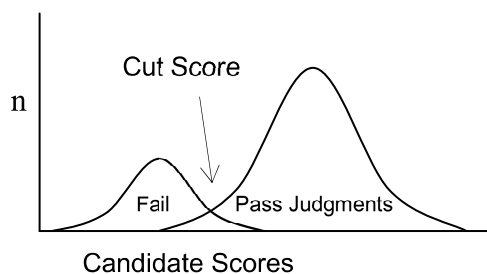
Requires judges to assess only the relevance/importance of each item and group them as essential, important or indicated. Judges then determine the number of items in each of the three categories that the borderline candidate should be able to answer correctly. The number correct is averaged across categories to determine the pass point.

Bookmark method

Item difficulties are determined from data from an actual exam administration. Items are arranged by item difficulty from least to most difficult. Judges select the most difficult item a borderline candidate would be likely to answer correctly and a “bookmark” is placed at that location. The passing scaled score for each judge's bookmark is calculated and performance data is then given to the judges. The impact of the collective average of pass points is discussed. The judges are then asked to place a second bookmark, which can be the same or different as the initial bookmark. A passing scaled score is calculated for each judge's second bookmark. The mean or median of the judges' passing scaled scores is the final cut score.

Contrasting Groups Method

A definition of what constitutes “qualified” is determined. The items are administered and scored. Groups of candidates with scores that represent “qualified” or “unqualified” are rescored in two separate distributions. The point at which the two distributions intersect is the location of the passing score.



Compromise Methods

Judges review a copy of the exam and determine the following information regarding the exam:

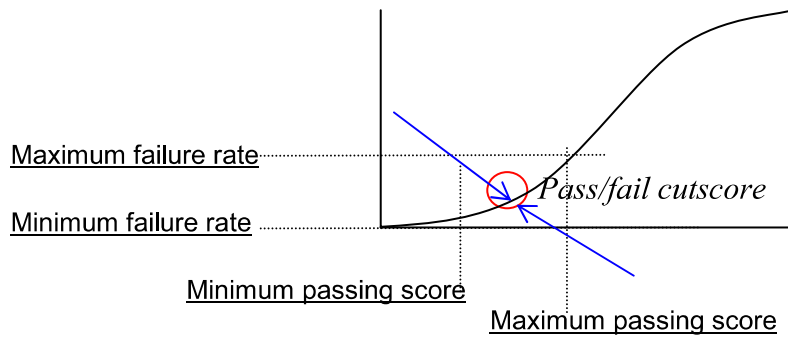
- Lowest acceptable percentage of failing candidates (minimum failure rate)
- Highest acceptable percentage of failing candidates (maximum failure rate)
- Lowest percent correct which would allow someone to pass (minimum passing score)

Highest percent correct required for someone to pass (maximum passing score)

The expectations of the judges are averaged to determine the four values that will be used. Those four values are plotted to create a rectangle (see image below).

After administration of the exam, the candidates' performance scores are plotted on the X axis (number correct) and Y axis (number wrong) resulting in a curve of candidate percent correct scores or ability estimates.

A diagonal line is drawn from the upper left to lower right of the rectangle. The point where the diagonal line intersects the plotted curve is the pass point.



A training session and/or discussion of the borderline candidates' knowledge, skills, and abilities are a key component to the psychometric quality of most standard setting methods.