



ASSESSMENT STRATEGIES INC.
CANADA'S TESTING COMPANY

STRATÉGIES EN ÉVALUATION INC.
LA SOCIÉTÉ SPÉCIALISTE DES EXAMENS AU CANADA

THE ANGOFF METHOD OF STANDARD SETTING

May 2014

STANDARD SETTING

Many agencies require credentialing for their professionals as one means of assuring the quality of practice. As a standardized examination is often a requirement for credentialing, determination of an appropriate pass mark for the examination is essential to the effectiveness of the process.

The Angoff method allows expert judges to determine an appropriate pass mark for an examination. A major advantage to this methodology is that the determined pass mark is based on the content of the examination and not on group performance.

The task for Angoff setting is usually assigned to members of the examination committee and is included as part the committee members' terms of reference. The task is typically conducted at the end of the examination committee meeting, following the approval of the new examination.

RELEVANT ISSUES

Setting a pass mark for an examination is setting a standard of performance on which decisions will be made about an individual's level of competence in a given field of practice. The pass mark determination is a judgement made by informed individuals (i.e., experts in the field of practice). It is determined through a rational discussion of the field of practice as well as an awareness of the consequences involved when a decision affecting individuals is made.

THE PASS MARK AND CONSEQUENCES

Whenever a pass mark is determined for an examination, there are a number of potential consequences that must be anticipated: an inappropriately low pass mark will allow non-competent candidates to practise, perhaps at the expense of the public welfare; an unrealistically high pass mark will exclude competent candidates from being credentialed.

The accuracy and precision of the measuring instrument (i.e., examination validity and reliability) must also be considered. Examinations are not perfect: they cannot include all the knowledge and skills in a given field of practice. An examination can only sample the field. Furthermore, if it were possible to repeatedly administer the same examination to a single candidate 100 times, the candidate's score would likely not be exactly the same each time. The inconsistency of the scores is a result of the reliability of the examination and the variables affecting candidate performance (such as anxiety level and health).

THE ANGOFF METHOD

The Angoff method requires expert judges to discuss the issues involved in determining a pass mark and to evaluate the examination by using a well-defined and rational procedure.

1. Competence and the Borderline Candidate

The Angoff method is based on the concept of the borderline or minimally competent candidate. The minimally competent candidate can be conceptualized as the candidate possessing the minimum level of knowledge and skills necessary to perform at a registration/licensure level. This candidate performs at a level “on the borderline” between acceptable and unacceptable performance. It is essential that each judge arrive at a clear and specific definition of the minimally competent candidate.

To better understand the concept of the minimally competent candidate, it is often helpful to think of the people you work with everyday; a few of them are the “superstars” performing at a level well above the majority, while others perform rather poorly and perhaps should not be practising. Somewhere between these two extremes is the group that performs at the level of minimum competence. The borderline candidate belongs to the group that just qualifies for registration/licensure.

2. Rating the Items

The Angoff method requires the judges to independently rate each item in the examination in terms of the minimally competent candidate. For each item, each judge answers the question: “In your opinion, what percentage of minimally competent candidates will answer this item correctly?” Alternately phrased, “Given 100 minimally competent candidates, how many will answer this item correctly?” The judge then indicates the appropriate percentage on the rating form and proceeds with the next item.

One common error made when rating items is to base the rating on the average candidate or the exceptional candidate rather than on the borderline candidate. Another potential error involves the interpretation of the question asked for each item. The items are to be rated in terms of how many borderline candidates will answer the item correctly. In a large group of borderline candidates, only some may actually know the correct response. It should not be assumed that all borderline candidates will know the answer. Finally, although item statistics may be used to provide additional information to the judges, the ratings should not be based solely on these statistics; item statistics are calculated on the entire candidate population, not on the borderline group alone.

3. Determining the Pass Mark

Once all the judges have rated each item in the examination, the ratings are collated and tabulated. The ratings for any single item should be in agreement. By agreement it is meant that the ratings for an item must all be within a certain percentage range (e.g., a 30%-range). If the range of the ratings is greater than the specified range, the judges providing the extreme ratings are asked to explain why they rated the item in that fashion. The other judges should explain why they rated the item as they did. If a brief discussion of these reasons does not cause any of the judges to change their ratings, the ratings should be left as they are. Once the discussion has ended, the average rating is calculated for each item and then for the total examination. This results in a percentage value which is the percentage score expected to be achieved by the borderline candidate. In addition to the expert ratings, a variety of relevant data is carefully considered to ensure that the standard that examinees will be required to achieve is valid and fair. This can include information on the preparation of new graduates, data on the performance of examinees on previously administered examinations, and pertinent psychometric findings. Based on all of this information, a point is set on a measurement scale that represents the minimum acceptable standard.

As an illustration of the rating process, consider a fictitious application of the Angoff method, with a panel of six judges setting the pass mark on a four-item exam. Following the orientation session, the judges provide independent ratings on each of the items in the exam. The ratings obtained are presented in Table 1. With a 30%-rule specified to define the target level of agreement, items 1, 2, and 4 require no post-rating discussion (i.e., for each of these items, the extreme ratings do not differ by more than 30%). Item 3, however, is identified as requiring discussion because of the 40% difference in the ratings of Judge 3 and Judge 5 on this item. As a result of the discussion, Judge 3 decreases her rating from 85% to 75% and Judge 5 increases his rating from 45% to 50%. The average ratings are then calculated for each item, and the average of these values is calculated, to arrive at an overall pass mark of 69%.

TABLE 1: Example of the Angoff Method

Item #	Judges' Ratings (%)						Average Rating
	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6	
1	65	70	65	65	70	65	67
2	85	70	60	80	70	80	74
3	70	65	85→75	70	45→50	70	67
4	75	60	65	70	75	70	69
Overall Average Rating							69

4. Factors for Successful Implementation

A number of factors contribute to the successful implementation of the Angoff method. An effective training session is essential in orienting the judges to the concept of the minimally competent candidate. As well, discussion and modification of extreme ratings help ensure that a defensible and valid cutoff score is established.

SUMMARY

The Angoff method allows expert judges to determine an appropriate pass mark for an examination, based on a discussion of the issues involved in credentialing and their assessment of the examination. A major advantage to this methodology is that the determined pass mark is based on the content of the examination and not on group performance. The pass mark is set in direct reference to the candidates, the competence required of the candidates, and the difficulty of the items themselves and, as a result, it is considered fair and valid.