

Considerations in Setting Cut Scores

By Roberta N. Chinn

COPYRIGHT 2006. Rights to copy and distribute this publication are hereby granted to members of the Council on Licensure, Enforcement and Regulation (CLEAR), providing credit is given to CLEAR and copies are not distributed for profit.

INTRODUCTION

Setting the cut score (standard setting) is the most important and the most troublesome step of any examination development process. The standard setting process is different from other psychometric processes with respect to its integrity, particularly for small programs, as it is heavily dependent upon expert judgment. The standard setting process is further complicated by the fact that in licensing and certification, there are no external criteria or sources to validate the cut score.

Setting cut scores is a highly complex process that involves setting policy and the conversion of judgments to a scale (Reckase, 2001). A licensing or certification board has the responsibility to ensure that the cut score is set according to technical and professional standards. If a candidate's score equals or exceeds the cut score, a board can grant a license or credential, confident that candidates who achieve a particular score possess the minimum knowledge and experience necessary to carry out the duties of a given profession.

In the case of small examination programs, there is no psychometrically sound means to determine a cut score other than a criterion-referenced cut score study. Small programs may hear about large-scale credentialing programs using item response theory (IRT) methods for equating examinations and large-scale series of focus groups to establish the cut score. However, many examination programs, such as those that administer 50-100 examinations per year, do not have the resources or the candidate volumes to establish cut scores statistically or with multiple focus groups that require large time commitments. It is not uncommon for state or provincial programs to test 40 or fewer candidates per administration, in which case, statistics can provide only gross information about the performance of an examination and its items.

This paper assumes a focus group approach and offers suggestions for preparing for and conducting cut score studies that result in valid cut score outcomes.

METHODOLOGY

There are a number of standard setting methodologies that may be used. The choice of methodology depends largely upon how the resultant cut score will be used. Whatever

method is used, the cut score must be acceptable to the board when applied to its examination results.

The ideal methodology focuses on criterion-referenced standards and use of focus groups of well-chosen experts. There are several traditional methods that could be adapted for use in the absence of stable statistics. All of the methods mentioned in this paper rely on the use of a definition of minimum competence as the criteria for determining the cut score.

Modified Angoff method. The most frequently used adaptation of the Angoff method (Angoff, 1971) involves asking judges to discuss the characteristics of a “borderline examinee.” For each item, judges provide estimates of the percentage of borderline examinees that would answer the item correctly. The cut score is based upon the average of the percentages for each item. This is a variation of the original methodology described by Angoff in which judges could envision a hypothetical person who was minimally competent and decide whether or not such a person could answer each item under consideration (yes-no).

Nedelsky’s method. The thrust of this method (Nedelsky, 1954) conceptualizes persons on the borderline between passing and failing and identifies the options in the items that persons would be able to reject as obviously incorrect. The reciprocal of the number of remaining options becomes the probability that the borderline person would answer the item correctly.

Ebel’s method. This method involves finding the point on the score scale “midway between the ideal mean score and the expected chance score” (Ebel, 1972, p. 492). Judges discuss the characteristics of the “borderline examinee” and rate the items based on difficulty (easy, medium, hard) and relevance (essential, important, acceptable, questionable). Judges then indicate the number of items in each category that the borderline examinee would answer correctly. The cut score is the percentage of possible points that a borderline examinee would obtain.

Beuk’s method. Beuk (1984) assumes that each judge has an opinion of what the cut score should be and what pass rate would be expected. In this method, each judge is asked to specify the percentage of the total raw score on the test, or minimum level of knowledge, that a person should possess in order to pass the examination. Each judge is also asked to specify the percentage of candidates that he or she expects to pass the examination. Standard deviations are measured for the judges’ percentage of the total raw score and the expected pass rate. The cut score is adjusted proportionally to the ratio of the standard deviations of the expected pass rate and the estimated percentage of the total raw score.

PRACTICAL ISSUES

Regardless of cut score method, several practical issues should be considered in planning a cut score study. These issues have a direct effect on the acceptance of and the validity of the cut score.

Social and political consequences. Even when it is done well, stakeholders, educational institutions, and professional organizations may have the perception that the cut score is arbitrarily derived. A board can be assured that the cut score process is psychometrically defensible so long as the examination is well constructed and the process is carefully conducted and well documented.

Statistical adjustments can be made to the cut score; however there is a danger of shifting the cut score too high or too low without psychometric rationale. The conservative approach is to conduct another cut score study or average the results from multiple cut score studies of the same examination.

Intended use of cut score. The general purpose served by the cut score should be defined, at least in general terms, prior to the cut score process. There must be mutual understanding between the board and the psychometrician as to how the cut score is to be used. Both board members and examination committees should have a clear idea about the meaning of the cut score and how it relates to minimum competence and protection of the public health, safety and welfare. There may be misconceptions (e.g., the cut score is an indication of competence required for employment or advanced credentialing; minimum competence sets a low standard of practice) regarding the meaning of minimum competence and how it relates to the cut score.

There must be a clear understanding that the cut score is the minimum score that candidates must attain in order to pass the examination so that committee members and/or focus group participants involved in the cut score process adopt the same perspective and use the same criteria for their ratings.

Cognitive demands of cut score process. The cognitive demands of the cut score process should not be underestimated. Cut score studies involve intense concentration and requires participants to understand many concepts and procedures quickly, e. g., prerequisite knowledge base and/or experience, definitions of minimum competence, subject matter of items, relative difficulty of items, cut score process, etc.

Qualifications of participants. Raymond and Reid (2001) note several factors to consider when identifying participants for a cut score study. These include the participants' ability to appreciate consequences of applying a standard, experience with the candidate population, knowledge of the levels of proficiency in candidate population, knowledge of the subject matter, analytical skills, and ability to concentrate for long periods of time. When composing focus groups, it is wise to consider representation from stakeholder groups, e.g. members of the examination committee, licensees who practice in specialties, etc.

Number of participants. The number of participants is dependent upon the number of stakeholder groups of the population. If the population of interest is fairly homogenous, then it is reasonable to convene a group of 6-8 participants. On the other hand, if the population is heterogeneous, a group might be as large as 10-12 participants. The participants should be representative of the population in terms of gender, age, experience, work setting, geographical region, etc.

Sufficient resources. Cut score studies represent the culmination of an entire examination development process and require monetary as well as staff resources. It is strongly recommended that sufficient resources be allocated for participants' travel and accommodations, preparation of training materials, professionals trained in psychometrics, etc.

FACTORS THAT IMPACT CUT SCORE OUTCOMES

There are a number of factors, many of them related to the items in the examination, which should be considered together with the practical issues in planning a cut score study.

Quality of the items. It cannot be overemphasized that the quality of the items has a tremendous impact on the rating process. Items should be in pristine condition before being included on an examination. There are several procedures that can ensure the quality of the items. Formal training of the item writers is essential for the construction of quality items.

Formal review of the items should be conducted by a focus group solely for the purpose of critically reviewing the items to evaluate overall content, readability, complexity, and amount of information presented. Throughout the review process, it is wise for reviewers to consider the perspective of a minimally competent candidate. The review process may be enhanced by including persons who have not been involved in the item writing process so that objectivity is maintained. Reviewers should be able to reach consensus regarding the authoritative reference and the correct answer.

The effect of thorough item review is a quality examination. Examinations that are well constructed will allow the discussion in a cut score procedure to focus on difficulty rather than the technical flaws of the items. This is particularly true if there are multiple rounds of rating, or if the item involves several computations or essay responses.

Cognitive demands required by the items. Haladyna (2004) classifies knowledge content into four categories (facts, concepts, principles, procedures) that involve two cognitive processes (recall, understanding). Items involving recall of facts, concepts, principles, or procedures are likely to require less time than items involving understanding of the same concepts, principles, or procedures due to the extent of cognitive demands imposed on the candidate.

Therefore, if an item contains tables of information, a list of existing conditions, and/or clinical or field test results, candidates are likely to take more time to answer the item, even in multiple-choice formats, because there are judgments to be made regarding relevance of information to the resultant diagnostics, computations, recommendations, etc. For example, an item based on verbatim reading of state law will impose different cognitive demands than an item based on the premise of a scenario and an interpretation of state law.

Item formats used. The physical structure of the item is somewhat related to, but distinct from, cognitive demands. There may be questions that are formatted as straightforward interrogative questions that ask candidates to identify concrete facts, concepts and procedures, apply rules, or construct diagrams. Even if the question addresses difficult concepts and procedures, candidates may perceive straightforward interrogative questions (e.g., what, which of the following, why) as “easier to understand.”

Other questions may be formatted as a three or four sentence scenario ending with an interrogative question. Candidates are asked to engage in a two-step process: first, perform an analysis of abstract constructs or case details in terms of prescribed procedures, computations or estimations, and then provide professional recommendations or predictions relative to their analysis.

Time allotted for the cut score study. It may take 1-2 days to conduct a cut score study. Cut score ratings will proceed more quickly for items that require knowledge of facts and principles whereas ratings will proceed more slowly for items that require analysis of case data, computations, diagnostic formulations or professional recommendations.

Criteria relating the minimally competent candidate. Development of criteria relating to minimum competence (“behavioral descriptors” or “performance level descriptors”) is an important step for setting performance standards in cut score studies (Mills & Jaeger, 1998; Hambleton, 1998).

Ideally, the criteria are developed prior to the cut score study by an independent focus group. There are many advantages to developing the criteria for the cut score study in advance of the meeting. First, the criteria will be detailed and cover each content area of the test specifications. Second, the criteria can be developed independently without encroaching on the tasks to be accomplished in the cut score study. Third, the criteria can be used for several cut score studies to maintain consistency of standards on which difficulty ratings are based.

The methodology for developing behavioral descriptors is derived from Flanagan’s “Critical Incident Technique” (1954). The exercise begins by asking practitioners to develop examples of how a person “would” perform a particular job activity in a specific context. The assumption of what a candidate “would” do is generally different than what a candidate “should” do in a given circumstance. The exercise includes developing descriptors for three performance levels, even though the minimal acceptable

competence level is used for the cut score study. The purpose of developing three performance levels is to assist the participants in identifying behaviors that define minimum competence in the context of other levels of performance.

For example, there are three levels of performance by which practitioners perform a given task. In the task “Interpret results of field tests,” a highly effective person might base an interpretation on multiple samplings from several field tests. A minimally competent person might base an interpretation on several samplings of a single field test. An ineffective person might base an interpretation on field logs that may not be sensitive to the substances of interest.

Oftentimes, participants view minimum competence in terms of the practitioners that they would never hire, or the practitioner that performed marginally in their academic preparation and training. In the development of competence criteria, participants may need assurance that minimum competence as related to cut scores refers to the level of competence acceptable to the profession such that the actions of a qualified practitioner will not harm the public.

The next step is to have participants review the tasks and knowledge involved in each content area of the test specifications. During this phase, the participants are given an opportunity to discuss the key points regarding what is covered in each content area. This step is critical to linking the behavioral descriptors to the framework of the examination.

Then, with the primary focus of the content area and associated job tasks fresh in their minds, participants are given an opportunity to generate behavioral descriptors for each content area. The descriptors are phrased in objective terms with regard to how a highly effective, a minimally competent, and an ineffective candidate would perform that task. Emphasis should be placed on the candidate who performs in a sophisticated manner that is “above and beyond” minimum competence (highly effective), the candidate who performs in an acceptable manner that does not harm the public (minimum competence), and the candidate whose behaviors would result in errors of omission or commission (ineffective). There may not be parallel descriptors for all job tasks; however, there are likely to be a number of descriptors that will apply to several job tasks.

The final step is to have participants review the list of behavioral descriptors and obtain consensus on their content so there is a common understanding of minimum competence.

A sample of behavioral descriptors is presented in Appendix A.

PREPARATIONS

There are several preparatory measures that can be completed prior to the cut score study that will enhance its success.

Selection of participants. Regardless of how they are chosen, it is very important that the participants in the focus group represent the communities of interest, have a good understanding of both the candidate population and the instructional environment, and appreciate the consequences of the cut score (Raymond & Reid, 2001). The bottom line is to have adequate representation of diverse practice settings and different levels of experience. Participants typically include a mix of practitioners who are educators or supervisors and practitioners who are newly licensed or certified so that many points of view can be represented in the final cut score.

Training materials sent to participants in advance of the process. The purpose of training materials is to provide an overview of the purpose and process that participants will be asked to use and the context of the ratings that they will provide. The materials will guide participants in their conceptualization of minimum competence and the impact of their ratings on the cut score.

Examination security. Participants at the cut score study should sign a nondisclosure or an examination security agreement in which individuals certify that specific information regarding the examination items or process are to be kept confidential, else, the entire examination is compromised.

Formal test publication. Oftentimes, new programs have relatively small item banks and few, if any, statistics for the items, and therefore, no empirical method by which to select questions for the examination. In these cases, the best way to select items for an examination is to convene a focus group for a one-day meeting and methodically select the highest quality items that meet the test specifications. The benefit of the focus group is to ensure that the items cover all aspects of the test specifications and the content of the items is independent of each other.

Final review of the examination. One might think that if all the checks and balances of the system were in place, that the final draft of the examination would be error-free. Theoretically, this is true; however, review of the final draft of the examination prior to the cut score study will ensure that participants in the upcoming cut score study are focusing their attentions on the difficulty of the items rather than technical inaccuracies, grammatical errors, or typographical errors.

The final review should include verification of all calculations, figures, factual data, and authoritative reference sources. The end result is a well-constructed examination that allows the substantive discussion of item content between rounds of ratings rather than discussion of the technical flaws of the items.

ELEMENTS OF A CUT SCORE STUDY

The key to establishing a defensible cut score with any cut score method is thorough training of participants. Thorough training is critical so that participants can understand what minimum competence is, what knowledge that a minimally competent candidate

brings to the examination, and what duties a minimally competent candidate can perform.

In the following example, we describe nine elements involved in establishing a cut score for a multiple-choice examination. The elements focus on cultivating a mental framework and mindset that participants can use to undertake the task of establishing the cut score during the course of a two-day cut score study. The examination could include items that require candidates to apply knowledge and training to straightforward factual items as well as items that require solutions, recommendations or predictions based on scenarios and tables of information.

It should be noted that the sample agenda in Appendix B uses a modified Angoff methodology but the principles could be adapted to other cut score methodologies or other types of examination formats such as performance examinations.

Element 1. Review the purpose of the examination.

Begin by having the participants consider the purpose of the examination. It is common for some participants to conceptualize the examination in terms of the ones that are given in an academic setting, and, other participants to conceptualize the examination in terms of a painful rite of passage. Both of these conceptions need to be eliminated prior to beginning the actual evaluation of the items.

It should be made clear to the participants that the purpose of a licensure or certification examination is to identify those candidates who possess the minimum competence to practice without harming the public health, safety, or welfare, rather than to classify persons into grade levels or identify the best job candidates for hire. The basic question that participants need to consider is, "Who is minimally qualified to practice?"

Element 2. Review the meaning of the cut score.

Have participants discuss the concept of minimum competence. Participants should be aware that the purpose of the cut score is to identify persons who possess the minimum qualifications to practice safely. They should also be aware that the cut score is essentially a pass/fail (dichotomous) decision. Therefore, with cut scores for licensing and certification examinations, participants must understand that high-scoring candidates do not necessarily know more than just-passing or low-scoring candidates. One way to explain the concept is in terms of a step function, e.g., "they (the candidates) either know it or they don't."

Element 3. Review minimum qualifications and test specifications for the examination.

It is very easy for participants with varying degrees of tenure to lose sight of what knowledge and experience that candidates currently possess before they take the examination. Oftentimes, participants are only familiar with their own level of knowledge and training before they took the examination or with the hiring criteria at their place of employment. But, as times change, veteran practitioners with 20 or more years of experience should become apprised of the current eligibility requirements for the examination.

In addition to the eligibility requirements to sit for the examination, it is recommended to review the content of major areas in the test specifications. By doing so, the participants can become familiar with the subject matter that is currently covered on the examination.

Element 4. Develop/review minimum acceptable competence criteria.

Although the list of “behavioral descriptors” or “performance level descriptors” could be developed prior to the meeting, they could also be developed on the same day as the cut score study. In either case, it is important to discuss the list of behavioral descriptors prior to the rating process so that participants understand the concept of minimum competence. Be sure to emphasize that minimum competence means the level of competence that would allow for protection of the public health, safety and welfare. Begin by introducing the descriptors as examples of how a person “would” perform a particular job activity. Then, discuss minimum competence in context with other levels of performance (highly effective, ineffective).

Element 5. Have participants take and self-score the examination.

The purpose of this exercise is to assist the participants in conceptualizing minimum competence rather than identify personal deficiencies. Participants may be skeptical of taking the examination; however, it is necessary for participants to envision what processes a candidate uses when taking the examination. The examination should be presented without the answers or solutions so that participants must focus on the knowledge and training that a minimally competent candidate must apply to the items. The correct answers are provided once all participants complete the examination so that the participants can score their own examinations. The psychometrician should reassure participants that their understanding of minimum competence, not their raw scores, is the key to the cut score process.

It should be emphasized that the participants' perceptions of the actual items on the examination are very important in rating the difficulty of the items. There may be specific topics addressed in the items that were not covered 10 years. There may also be more stringent eligibility requirements that must be fulfilled prior to taking the examination.

Element 6. Getting acquainted with the rating process.

Ideally, non-operational items are used to train participants in the rating process; however, it can be done with the first few operational items on the examination. Throughout the process, the psychometrician may have to remind participants not to work ahead of the group so that participants can focus their full attention on the final ratings on the merits of each item.

The rating process should include at least two rounds of ratings so that there is an opportunity for judges to review their decisions before the cut score is finalized (e.g., Busch and Jaeger, 1998; Hambleton, 2001). Initial ratings are made independent of the group for a fixed set of items and the final ratings are made in the presence of other participants. The process of providing opportunities for feedback is important in reducing the amount of variability in ratings and the amount of error in the ratings because judges have the opportunity to correct any misunderstandings related to item difficulty (Reckase, 2001).

For example, in a modified Angoff session, participants may be asked to make two rounds of ratings. The psychometrician should emphasize that the standard setting process is a group process, not an individual one. Initially, participants provide independent Round 1 ratings one item at a time for the first 5-10 items and receive feedback from other participants in terms of rationales for the ratings. Each participant would get a turn at providing a rationale for his/her rating before moving on to the next item.

Then, participants provide Round 2 ratings on items one at a time or until the psychometrician is confident that participants are astute to each other's feedback and to the minimum competence criteria. The Round 2 ratings provide an opportunity for participants to change their ratings in light of the feedback.

In summary, the procedure is carried out item by item, initially in blocks of 5-10 items, and later in blocks of 10-20 items. The psychometrician would ask each of the participants what their Round 1 rating was and what their rationale was for the first item. Then, after hearing all the discussion, participants make their Round 2 (final) ratings for the first item in light of the discussion. Participants proceed to the second, third, and fourth item, and so on until there are Round 2 ratings for all items.

Element 7. Monitoring participants' understanding of the cut score process.

The psychometrician must continually assess participants' ratings in such a way as to allow participants to provide their ratings yet monitor their ratings so that participants are not ignoring the concept of the minimum competence and are not promoting their personal opinions. Specifically, participants' ratings are monitored for stability over occasions, consistency with the concept of minimum competence, and consistency with realistic expectations. The psychometrician can track the ratings on a copy of the examination or on a dry-wipe board, intervening only if it is clear that participants are not following instructions.

Since there is no correct rating for an item, the psychometrician must decide when to intervene and how often. It is important for the psychometrician to be aware of the cumulative effect of intervention which may focus undue attention on the ratings of specific participants, e. g., those with very high or very low ratings. These participants may decide to "give in" to the group without having a justifiable reason for their ratings to avoid being singled out.

Element 8. Monitoring participants' use of process feedback between rounds.

A psychometrician has to unobtrusively monitor each participant's use of feedback throughout the cut score process. The cut score could be affected if one participant proceeds without incorporating the comments of other participants at the time his/her ratings are made. Every effort should be made to elicit responses from the participants such that Round 2 ratings are assigned after conclusion of discussion regarding the test items.

Element 9. Close the meeting and obtain feedback from participants.

It is important to provide opportunities for participants to provide feedback regarding the process. Not only can feedback provide information that may be helpful in interpreting the cut score ratings, the feedback can be helpful to improve future cut score studies.

The feedback may also include evaluation of the cut score process in the form of a short questionnaire. Such questionnaires may ask participants to mark "agree" or "disagree" to several statements (see Cizek, Bunch & Koons, 2004, p. 45) such as "The orientation provided me with a clear understanding of the purpose of the meeting," "The time provided for discussion was adequate," "I was able to follow the instructions and complete the rating sheets accurately."

SUMMARY

A valid cut score is dependent upon many considerations and should be undertaken with a clear understanding of minimum competence and the intended use of the cut score. There are a number of practical issues that affect the acceptance of and validity of the cut score including social and political consequences, the cognitive demands of the cut score process and the qualifications of participants. There are also factors that impact the cut score outcome, most notably, the quality of the items in the examination. A high-quality examination is the cornerstone of a cut score study because it allows participants to focus on the difficulty, rather than the technical flaws, of the items. When implementing a cut score study, participants should be thoroughly trained in the cut score process and their ratings monitored for adherence to minimum competence criteria. Thus, with careful planning, preparation and implementation, the result will be a valid cut score that is psychometrically defensible.

REFERENCES

- Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), Educational Measurement (pp. 508-600). Washington, DC: American Council on Education.
- Beuk, C. H. (1984). A method for reaching a compromise between absolute and relative standards in examinations. Journal of Educational Measurement, 21, 147-152.
- Cizek, G. J., Bunch, M. B., & Koons, H. (2004). Setting performance standards: Contemporary methods. Educational Measurement, 23, 31-50.
- Ebel, R. L. (1972). Essentials of educational measurement (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Flanagan, J. C. (1954). The critical incident technique. Psychological Bulletin, 51, 327-358.
- Haladyna, T. M. (2004). Developing and validating multiple-choice test items (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hambleton, R. K. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. In G. J. Cizek (Ed.), Setting performance standards: Concepts, method and perspectives. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mills, C. N., & Jaeger, R. M. (1998). Creating descriptions of desired student achievement when setting performance standards. In L. Hansche (Ed.), Handbook for the development of performance standards (pp. 73-85). Washington, DC: US Department of Education and the Council of Chief State School Officers.
- Nedelsky, L. (1954). Absolute grading standards for objective tests. Educational and Psychological Measurement, 14, 3-19.
- Raymond, M. R., & Reid, J. B. (2001). Who made thee a judge? Selecting and training participants for standard setting. In: G. J. Cizek (Ed.), Setting performance standards: Concepts methods and perspectives (pp. 119-157). Mahwah, NJ: Erlbaum.
- Reckase, M. D. (2001). Innovative methods for helping standard-setting participants to perform their task: The role of feedback regarding consistency, accuracy, and impact. In: G. J. Cizek (Ed.), Setting performance standards: Concepts methods and perspectives (pp. 159-173). Mahwah, NJ: Erlbaum.

APPENDIX A – Sample behavioral descriptors

HIGHLY EFFECTIVE	<ul style="list-style-type: none">• Applies multiple theoretical models to the case• Integrates information from medical, psychological, psychiatric, and other referral sources into treatment plan• Selects treatment model that provides optimum benefit to client• Establishes a termination plan that addresses termination issues
MINIMUM COMPETENCE	<ul style="list-style-type: none">• Demonstrates familiarity with basic concepts and elements of multiple theoretical models• Selects treatment models suited to specific information in the case• Prioritizes treatment issues when developing treatment plan• Modifies and revises treatment plan as indicated by ongoing assessment• Develops follow-up plan that client can implement• Engages client in treatment-planning process
INEFFECTIVE	<ul style="list-style-type: none">• Develops a treatment plan with no theoretical framework• Develops treatment plan based on inaccurate or incorrect interpretation of theoretical frameworks

APPENDIX B – Sample agenda for cut score study¹

DAY ONE

9:00 am – 9:20 am	Introductions
9:20 am – 9:30 am	Address topic of examination security
9:30 am – 10:00 am	Overview of cut score procedures
10:00 am – 10:30 am	Review purpose of the examination
10:30 am – 10:40 am	Review meaning of the cut score
10:40 am – 11:00 am	Review minimum qualifications and test specifications for the examination <ul style="list-style-type: none">• Requisite knowledge and experience• Content areas in the test specifications
11:00 am – 12:00 pm	Review minimum acceptable competence criteria
12:00 pm – 1:00 pm	Lunch ²
1:00 pm – 2:00 pm	Continue reviewing minimum acceptable competence criteria
2:00 pm – 4:45 pm	Take examination “as a candidate”
4:45 pm – 5:00 pm	Self-score examination

DAY TWO

9:00 am – 9:30 am	Discussion regarding minimum competence and rating process
9:30 am – 10:30 am	Orientation to cut score rating process (5-10 items) <ul style="list-style-type: none">• Round 1 – independent ratings• Each participant explains rationale for ratings• Round 2 – final ratings in light of feedback
10:30 am – 12:00 pm	Continue Round 1 and Round 2 ratings (blocks of 5-10 items)
12:00 pm – 1:00 pm	Lunch
1:00 pm – 4:00 pm	Continue Round 1 and Round 2 ratings (blocks of 5-10 items)
4:00 pm – 4:30 pm	Wrap-up and adjourn meeting

¹This agenda is designed for a modified Angoff, but could be adapted to other cut score methodologies or other types of examinations such as performance examinations.

² In addition to lunch, 10-15 minute breaks can be incorporated into the morning and afternoon sessions