APPENDIX G

Angoff Rating Instructions

Criterion-Referenced Cutoff Score Determination Subject Matter Expert Training Outline

CPS

Criterion-Referenced Cutoff Score Determination
Subject Matter Expert Training Outline

Introduction

- *The nature of passpoints. The discussion covers what a passpoint is and what it means.*
    - *Arbitrary--70%*
    - *Norm-referenced*
    - *Criterion-referenced*

Angoff Method

- *An overview of the modified Angoff method is discussed.*

Minimally Acceptable and Superior Candidates

- *Raters discuss what it means to be a minimally competent candidate versus a superior candidate. Emphasis is placed on understanding that minimum competence does not mean substandard.*

Development of Behavioral Descriptions of each Competency Level

- *Use Task/Occupational Analysis as a basis.*
- *Reach consensus amongst meeting participants on wording, meaning, and so on.*

Discussion on How to Assign Angoff Ratings

- *Additional instruction is provided to raters on how to assign Angoff ratings for each item.*

Calibration Session

- *To ensure raters are using the same thought processes when making their ratings, the raters are asked to evaluate a subset of test items. Next, raters are asked to report their ratings to the group. If there are significant differences between ratings, the raters are asked to explain how they arrived at their ratings. Consensus is then attained on the ratings. This process in repeated for each calibration item.*

Rater Monitoring

- *Throughout the rating session, raters are periodically asked to report their ratings for particular items to ensure consistency. Raters who appear to be group outliers receive additional individualized training.*

Criterion-Referenced Cutoff Score Determination

Background

In licensure and certification testing it is a common practice to use a criterion-referenced approach to set the passpoint (cutoff score) for an examination. This approach provides a defensible rationale for identifying a cutoff score. The cutoff score is estimated using the Angoff method.

The main rationale behind criterion-referenced cutoff scores is that one must be able to distinguish between candidates who can demonstrate sufficient knowledge to be licensed or certified and those who cannot. The cutoff score creates two classifications of candidates: those who are competent to perform the duties in a safe and effective/ definitive manner and those who may not perform at this level.

Defining levels of Competency

In order to distinguish between those who are competent to do the job and those who are not, we must first define levels of competency as they relate to observable work behaviors. There are two levels of competency that must be defined.

Minimal Acceptable Candidate (MAC): a concept that describes what the minimally acceptable candidate should be able to do, or should know, on the very first day on the job. The minimally acceptable candidate should be defined according to observable work behaviors through which a person can demonstrate his/her knowledge, skills, or abilities.

Superior Worker: a concept that describes what the superior, or highly qualified person should be able to do, or should know, on the very first day on the job. The superior worker should be defined according to observable work behaviors through which a person can demonstrate his/her knowledge, skills, or abilities.

The Angoff Method

The Angoff method defines the cutoff score as the lowest score the minimally acceptable candidate is likely to achieve. Candidates scoring below this level are believed to lack sufficient knowledge, skills, or abilities to be certified.

For each test item, subject matter experts are asked to estimate the probability that a minimally competent applicant will get the answer correct. These probabilities are then summed and divided by the number of subject matter experts. The result is the passing standard.

In general, the appropriate passing point will correspond to a judgement of the number of items that can be answered correctly by a minimally acceptable candidate. This judgement is made by subject matter experts who evaluate the test material on an item-by-item basis. The evaluation process is commonly called an Angoff Review, and is named for one of the psychologists who originally developed the technique.

A minimally acceptable candidate is one who meets the standard, though barely. A minimally acceptable candidate has enough of the requisite knowledge and skill to do the job, although their technical background may be limited. A minimally acceptable candidate is borderline, but acceptable.

In performing an Angoff Review, each subject matter expert is asked to assign a number to each test item, based on the percentage of minimally acceptable candidates who might reasonably be expected to answer it correctly. Thus, if all or nearly all minimally acceptable candidates would be able to answer the item correctly, it would be assigned a value of 95-100. However, because even strong candidates come from a variety of backgrounds and are not expected to know everything, many perfectly good items will receive lower numbers. With traditional four-choice items, the lowest number assigned would be 25, which represents the percentage of candidates who might be expected to get the item correct by guessing.

Angoff ratings by subject matter experts are pooled and averaged to yield an overall average for the test. This overall average is the expected total score for a minimally acceptable candidate and becomes the tentative passing point, expressed as a percentage of possible total points. After the test is administered, statistical analysis is performed on both the individual items and the test as a whole, and the final pass point determination is made. Usually, because of limitations of the test and other considerations, the results of the statistical analysis may suggest a passing point slightly below that derived from the Angoff Review.

INSTRUCTIONS FOR RATING TASKS

For each item you are asked to provide data on three separate dimensions. First you are asked to assign an Angoff rating for each item. The instructions are as follows:

> For each item, enter a number from 25 to 100 in the "Minimum Acceptable" column which represents your estimate of the percentage of minimally acceptable candidates who you believe would answer the item correctly. Your item rating can also be considered the probability that a minimally acceptable candidate would answer the item correctly. Thus, a rating of 50 would mean that a minimally acceptable candidate would have about a 50-50 chance of choosing the correct answer and a rating of 80 would mean an 80% chance, and so on. Brief notes can be made in the appropriate column, though these are not required. You should definitely make a note, however, if you believe the item is incorrectly keyed or is otherwise seriously defective.

> First look at the item without consulting the key. Look closely at the wrong answers as well as the right one. If a minimally acceptable candidate could eliminate one or two alternatives then the probability of the minimally acceptable candidate choosing the right answer is increased. For example, the minimally acceptable candidate who can eliminate two wrong answers gets to select from only two alternatives and hence has at least a 50% chance of picking the right

answer. Obviously, if candidates can easily eliminate all the wrong answers, they will all get the item correct. On the other hand, remember that minimally acceptable candidates are borderline candidates whose knowledge and skills, while acceptable, are limited, and that not even strong candidates are expected to earn perfect scores**.**

For the second set of ratings, you are asked to determine how appropriate an item is as a measure of the three disciplines (i.e., Firearms, Toolmark, and Gunshot Residue). If an item appears appropriate for the discipline circle "Yes." If not, please circle "No." Bear in mind that an item may be relevant for more than one discipline.

Finally, in the "Comments" section note any problems or areas of concern with a particular item. For example, an item may be measuring information that is only tangentially related to the discipline.

Please be sure to put your name on each page.