Cut Score Study Methodology

The judges serving on the standard setting study panel were selected by ABTC, all of whom were deemed to possess subject matter expertise. They were selected to provide for an appropriate balance on potentially relevant professional characteristics, such as area of special expertise, practice setting, and geographical distribution.

The judges participated in a standard setting study that consists of the following three major steps:

1. Definition of Minimum Competence

2. Rating of Examination Items

3. Consideration of Empirical Data

Definition of Minimum Competence

In preparation of the rating process, a discussion regarding the definition of a minimally competent practitioner (MCP) was facilitated. An MCP is described as an individual who has enough knowledge to practice safely and competently but does not demonstrate the knowledge level to be considered an expert.

Rating of Examination Items

Judges were then trained on the rating process. Central to this process is the notion that each rating is provided individually by each rater and reflects the answer to this question: What percentage of MCPs do you expect will answer this item correctly?

The judges were shown each item one at a time and instructed to provide a rating for each item (round 1 rating) after reading the stem and the response options. Judges then recorded their ratings in a spreadsheet.

Consideration of Empirical Data

Following the initial rating for each item, the answer key and p-value were presented so that the judges could re-evaluate their thought process and revise their ratings (round 2 rating). For example, the judges were specifically advised to consider the possibility that their ratings might be too high on items that they answered incorrectly when the initial ratings were recorded or if their expectations of performance for MCPs were significantly different from the p-value, which represents the performance of a sample of

examinees who represent all levels of competency and performance. Judges then recorded their round 2 ratings in the same spreadsheet.

Ratings were then collected from the raters, and items for which the average rating was higher than the p-value by 5 points or more were identified as well as those in which the highest and lowest ratings differed by 40 points or more. The judges discussed these items to determine why their expectations differed significantly from the difficulty for all examinees or from each other. All raters were then given the opportunity to revise their ratings (round 3 rating) for the subset of items identified, following this discussion. Judges called out their round 3 ratings for this subset of items and the facilitator entered those ratings on screen.