**CR Values**

The ABMM uses a criterion-referencing system (the modified Nedelsky method) to assign each item (question) a difficulty value. This value is determined by how difficult the Exam Development Committee perceives an item to be, which corresponds directly to the number of sophisticated distractors within the item.

Here is a simple example:

Which body of the United States government has the power to declare war on another country?

A. The Executive Branch (-1)  
B. The Congress (1)  
C. The Supreme Court (-2)  
D. The Federation of States (-2)

CR Value = 0.60; there is one sophisticated distractor.

<table>
<thead>
<tr>
<th>KEY: Sophisticated Distractors</th>
</tr>
</thead>
</table>
| 1 | Correct answer  
This is the only correct answer. |
| -1 | Sophisticated distractor  
Examinees with *some knowledge* of the subject *might* choose this response. |
| -2 | Non-sophisticated distractor  
Examinees with *minimum knowledge* of the subject *would not* choose this response. |

<table>
<thead>
<tr>
<th>KEY: CR Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty Value</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>.90</td>
</tr>
<tr>
<td>.60</td>
</tr>
<tr>
<td>.45</td>
</tr>
<tr>
<td>.36</td>
</tr>
</tbody>
</table>

* This is the expected candidate correct response rate for a group of minimally competent examinees (i.e., those with just enough knowledge to competently perform the job being assessed by the exam). Actual candidate response rates may differ, depending on the strength (or weakness) of the candidates in a given year.
**DISCRIMINATION**

Discrimination is a function of how the highest scoring examinees did in comparison to the lowest scoring examinees. This index can range from -1.00 (weak examinees significantly outperform strong examinees on the item) to +1.00 (strong examinees significantly outperform weak examinees on the item). In other words, a question’s discrimination will be positive if the stronger examinees scored better on that item than the weaker examinees. The discrimination values are also applied to the distractors (i.e., incorrect answers).

Usually, positive discrimination above +0.20 for a correct answer, and low or negative discrimination (i.e., below +0.20) for the distractors, is a sign of a good item. An item showing a low or negative discrimination for the correct answer indicates that lower scorers on the exam scored almost as well or better on that item than the higher scorers did. Similarly, an item showing a positive discrimination for a distractor indicates that higher scorers on the exam were attracted to that distractor at a greater rate than the lower scorers. A discrimination value of zero (0.00) indicates that weak and strong examinees performed equally well.

**CUT (PASSING) SCORE**

The cut (pass/fail) score of an exam is directly related to the number of easy, medium, and hard questions appearing on that exam. For example, if all of the items on an exam have CR values of 0.90 (i.e., they are all easy questions), examinees will need to answer 90% of the items correctly to pass.

The ABMM uses an assessment software tool to generate exams each year. The software pulls questions from the pool that meet the exam’s content requirements and the average degree of difficulty (which has been set at 0.70 since 1999). The average degree of difficulty ensures the rigor of the exam is consistent from year-to-year.

Following the administration of the exam, all items appearing on the exam are reviewed. Their performance is compared to their CR and discrimination values to assess which items have performed as expected and which items should be reviewed by the ABMM Chair, Vice-Chair, and Exam Development Chair.

In the case of items that do not perform as expected, the current year’s performance statistics are compared to those of previous exams, if the data is available, to discern whether any discrepancies are due to an anomaly in examinee knowledge. If previous statistics are unavailable, the ABMM Chairs review items to ensure that they are not ambiguous or incorrect.

Items with content flaws are removed from scoring, and the cut score is calculated based on the average difficulty of the items remaining on the exam. The exam is not scored on a curve; each examinee’s score is derived solely from the number of questions answered correctly.