

National Board of Medical Examiners

Customized Examination Score Interpretation Guidelines

NBME customized examinations provide medical schools with a tool for measuring students' understanding of a series of content areas defined by each school. Through this process, schools can create customized examinations targeted to meet the specifics of a given curriculum. This document provides guidance for interpreting the feedback obtained from the customized assessments. Importantly, NBME neither sets nor recommends a "passing" score. Generally, customized examinations should be used in conjunction with other indicators of student performance to determine grades.

Interpreting Scores

Percent Correct Scores: Percent correct scores are computed for the total test and the content areas defined during the test construction process. Percent correct scores represent the percentage of items answered correctly within the corresponding content area. Percent correct scores are influenced by the overall difficulty of the exam; therefore, these scores can only be used to compare performance on test administrations of the same customized exam form. Percent correct scores should not be interpreted as percentile ranks.

Scaled Scores: Scaled scores are computed when the scaling group has at least 25 examinees. When this requirement is met, scaled scores are computed for the total test and the content areas that were defined during the test construction process. Scaled scores are computed to have a mean of 70 and a standard deviation of 8 for the scaling group of examinees identified in the heading of the scale score performance report. Scaled scores can be used to compare performance of examinees within the scaling group across different test administrations of the same exam form. Thus, scaled scores provide a useful tool for comparing the scores within the same administration as well as comparing scores across test administrations that used the same scaling group. Though the scaled scores may appear similar to percent correct scores, they are not comparable.

It is important to note that the scaled scores produced for an administration can only be compared to other administrations of the same exam form that used the same scaling group.

Performance Reports

Percent Correct Rosters: A Roster of Percent Correct Scores is provided for each administration. The roster of scores provides percent correct scores for all examinees on the total test and each content area selected to appear on the roster. A content area must have at least 20 items in order for the numerical score to be reported due to low reliability when content areas contain insufficient items. All reported scores also appear in an excel (XLS) file that can be downloaded and used to import scores into your local database.

A frequency distribution of the total test percent correct score for the total group is provided if two or more examinees tested during the administration. The distribution shows the number and percentage of examinees at each score, along with corresponding cumulative frequencies and percentages. Group level summary statistics are also provided for the total test and content areas.



National Board of Medical Examiners

Customized Examination Score Interpretation Guidelines

Scaling Group: For the first administration of an exam, the scaling group consists of examinees from that administration only. For subsequent administrations of the same exam, your school has the option to select the scaling group to be used to compute scaled scores. Scaling groups for the same exam can be defined based on examinees from the current administration, previous administrations, or a combination of both. When a school selects the same scaling group for all administrations of an exam, it allows examinee scaled score performance from these administrations to be compared.

In order to receive scaled score feedback, which includes a roster and individual performance profiles, you must define a scaling group of at least 25 examinees and link it to the administration; otherwise, only a percent correct score report will be posted. Please access the Scaling Group Application on the NBME Medical School Portal to define the scaling group. Note that examinees who did not take the exam under standard timing conditions, scored more than 3 standard deviations below the mean, or omitted more than 10% of the items are excluded from the scaling group.

Scaled Score Rosters: The Roster of Scaled Scores displays a scaled score for all examinee on the total test and each content area selected to appear on the roster. A content area must have at least 20 items in order for the numerical score to be reported on the roster due to low reliability when the content area contains insufficient items. All reported scores also appear in an Excel (.XLS) file that can be downloaded and used to import scores into your local database.

A frequency distribution of the total test scaled score for the total group is provided if two or more examinees tested during the dates shown above. The distribution shows the number and percentage of examinees at each score, along with corresponding cumulative frequencies and percentages. Group level summary statistics are also provided for the total test and content areas.

Individual Performance Profile: To aid identifying relative strengths and weaknesses, individual performance profiles reporting total test scaled and percent correct scores as well as performance bands are provided if there were at least 25 examinees in the scaling group. Profile bands display the range of plausible performance on the total test and content areas selected by a school. Due to increased measurement error for content areas with few items, a content area must have at least 10 items to appear on a profile. Profiles are posted as separate report files in a zip file, along with a consolidated report file that contains each examinee's profile sorted alphabetically by last name.

Item Analysis Report: A PDF and Online Item Analysis (IA) report is provided to help interpret performance on each item if 10 or more examinees within an administration met the criteria for inclusion in the scaling group defined above. The IA report presents content area information and item difficulty statistics. The Online IA also contains item text and images, along with the percentage of examinees selecting each option. If the IA at least 40 examinees meet the IA criteria, item discrimination indices are provided. All school level calculations are based on the examinees meeting the scaling group criteria. Historical item difficulties for each item based on the USMLE Step examinations, and/or an NBME Subject Examination are provided for gross comparative purposes. Since these values were computed when items appeared in examinations administered in different contexts, caution is advised when interpreting and comparing historical statistics. Most of the statistical item information is also provided in the keyword Excel (.XLS) file.

Further guidance on interpreting the Online IA report can be viewed by clicking on the Help link on the IA screen.