

THE TESTING COLUMN

MBE “DECIMAL DUST”

by Susan M. Case, Ph.D.

We have devoted previous column space to discussions of the difference between raw and scaled scores, because examinees (and others) have found the two scores to be confusing. As most of you know, the raw score is a simple count of the number of questions answered correctly. If an examinee answered 117 questions correctly, his raw score would be 117. In the case of the MBE, it would mean that the examinee answered 117 out of 200 questions correctly. The raw score is not very useful because comparisons of raw scores between varying examinations are not meaningful. Someone who answered 118 questions correctly on one exam might be more or less able than someone who answered 117 questions correctly on a different exam. While this might seem like a minor difference, misinterpretation of someone's performance because of a one-point shift can be important to examinees with scores near the pass/fail point.

The scaled score is the more important score. The scaled score reflects an adjustment to account for the change in difficulty from one exam to the next. For example, examinees who take different examinations and who each receive a scaled score of 140, performed at the same level, regardless of the number of questions answered correctly.



The confusion between raw and scaled scores is a common problem across the testing industry. Many testing programs report only scaled scores, but NCBE has historically provided both scores.

Let me digress for a moment to discuss score scales in general. Score scales are created; they are somewhat arbitrary. Fahrenheit and Celsius are examples of scales we use to define temperature in the everyday world. In the world of examinations, test developers try to establish a score scale that will facilitate interpretation rather than hinder it; they go to considerable lengths to establish a scale that will accurately reflect differences among scores without overstating these differences. They also attempt to avoid scales that are likely to be confusing.

The first decision is how broad to make the score scale. The advantage of a broad range is that each additional raw score point will produce an increase in the scaled score. The disadvantage of a broad range is that small differences in the raw score could appear to be larger than they are. For example, if the MBE scale ranged from 1 to 400, on average each increase of one raw score point would result in an increase of two scaled score points. This sort of broad scale magnifies the differences between scores. A difference of 15 scale points, for example, would seem more important than it is. At the other extreme, we could consider a very restricted score scale. With 200

questions on the MBE, if the scale ranged only from 1 to 10, for example, several raw scores would map into the same scale score—a scale score of ‘5’ might be awarded to examinees who answered as few as 100 or as many as 120 questions correctly.


The second decision is where to start and stop the scale. A common problem is using a scale that includes the values 1 to 100, which can be misinterpreted to be the percentage of questions answered correctly or the percentile rank of the examinee, neither of which would be the correct interpretation. The SAT scale was developed to range from 200 to 800 to avoid this problem. The score scale for the USMLE (United States Medical Licensing Exam) was developed to range from 120 to 280. Regardless of the care used in creating a score scale, it is almost inevitable that whatever scale is chosen, some confusion will result.

The MBE scaled score was created to look very similar to the raw score scale. While this approach has advantages, it has at least two disadvantages. First, as noted above, the raw score and the scaled score can be confused. The second disadvantage is that, at times, two different raw scores may yield the same scaled score due to rounding. For example, if someone answered 130 items correctly, he might receive a scaled score of 137.8, which rounds to 138. A second person who answered 131 items correctly might receive a scaled score of 138.4, which also rounds to 138.

Such disadvantages have led us to discuss whether the MBE score scale should be changed. However, changing the score scale for the MBE (or any other existing testing program) is a major undertaking. After deciding what the new score scale will look like, the change must be implemented in a way that is as painless as possible for the various constituencies. The transition to any new scale would be awkward. Depending on the characteristics of the

new scale, examinees who failed the MBE in the past might have a score history of 115 and 120, followed by a 65 on the new scale; or a 115 and 120, followed by a 180 on an alternative new scale. At the very least, it would be confusing.

After considerable discussion about the issues that surround the MBE scaled score, we ultimately decided to maintain the current score scale, but to offer it in two formats. The first format is exactly the same as the one currently used. The second format is otherwise identical, but scaled scores will be reported to the first decimal place (e.g., 137.6). The addition of a single decimal place will clarify which score is the raw score and which is the scaled score. In addition, the decimal place will ensure that each unique raw score is represented by a unique scaled score. Third, the new scale will not magnify the differences between scores (a 141.6 appears to be slightly higher than a 140.8, which is an appropriate interpretation). And, finally, the transition from the old scale to the new one should not be unduly awkward. Scores from the past will have the same meaning and can be interpreted correctly on the new scale.

We expect to provide the new decimal format for the July 2004 examination results. Each jurisdiction will have the choice of receiving the old form of the scale score without a decimal place or the new scale score with a decimal place. For the sake of consistency, we hope jurisdictions will all opt for the decimal version. We believe the format with one decimal place will help to clarify the difference between the raw score and the scaled score; we also believe it is at an appropriate level of precision—neither too coarse nor too fine. For more discussion of this point on precision, see the article by Mike Kane in the last issue of *THE BAR EXAMINER*, November 2003, pp. 24-29. 

SUSAN M. CASE, PH.D., is the Director of Testing for the National Conference of Bar Examiners.