

THE TESTING COLUMN

FREQUENTLY ASKED QUESTIONS ABOUT SCALING WRITTEN TEST SCORES TO THE MBE

by Susan M. Case, Ph.D.

Scaling is a topic that often arises at NCBE seminars and other meetings and has been addressed in *THE BAR EXAMINER*; yet it still seems to be a mysterious topic to many. This column addresses the most frequently asked questions about scaling in an attempt to clarify some of the issues.



The equating process requires that a mini-test comprised of items that have appeared on earlier versions of the test be embedded in the larger exam. The mini-test mirrors the full exam in terms of content and statistical properties of the items. The repeated items provide a direct link between the current form of the exam and previous forms. They make it possible to compare the performance of this administration's candi-

What is scaling?

In the bar examination setting, scaling is a statistical procedure that puts essay or performance test scores on the same score scale as the Multistate Bar Examination. Despite the change in scale, the rank ordering of individuals remains the same as it was on the original scale.

What is the outcome for bar examiners who do not scale their written test scores to the MBE?

To understand the effect of not scaling written scores to the MBE but keeping them on separate scales, one must consider the equating process that adjusts MBE "raw" scores to MBE "scaled" scores. As you know, equating ensures that MBE scores retain the same meaning over time, regardless of the difficulty of the test form that a particular examinee took and regardless of the relative proficiency of the pool of candidates in which a particular examinee tested.

dates with the performance of a previous group on exactly the same set of items. Even though different forms of the MBE are designed to be as similar as possible, slight variations in difficulty are unavoidable. Similarly, candidate groups differ in proficiency from one administration to the next, and from one year to the next. Equating methods are used to adjust MBE scores to account for these differences, so that a scaled score of 135 on the MBE in July 2004 represents the same level of proficiency as a scaled score of 135 on the MBE in February 2007 or on any other test.

Equating is not possible for written tests because written questions are not reused. As a consequence, essay scores will fluctuate in meaning from administration to administration because it is impossible for graders to account for differences in the difficulty of the questions or for differences in the average proficiency of candidates over time. This phenomenon is

demonstrated by the fact that average essay scores in February tend to be the same as average essay scores in July, even though we know that February candidates are consistently less proficient (as a group) than July candidates. It has also been shown that an essay of average proficiency will be graded lower if it appears in a pool of excellent essays than if it appears in a pool of poor essays. Context matters.

So what is the outcome of such fluctuation in the meaning of written test scores? An individual of average proficiency may have the misfortune of sitting for the bar with a particularly bright candidate pool. This average individual's essay scores will be lower than they would have been in a different sitting. The same individual's MBE score will reflect his genuine proficiency level (despite sitting with a group of particularly bright candidates), but without scaling, his essay scores may drag him down. An unscaled essay score may be affected by factors such as item difficulty or the average proficiency of the candidate pool that do not reflect the individual candidate's performance.

What is the outcome for bar examiners who scale their written test scores to the MBE?

The preferred approach is to scale written test scores to the MBE—this process transforms each raw written score into a scaled written score. A scaled written test score reflects a type of secondary equating that adjusts the “raw” written test score after taking into account the average proficiency of the candidate pool identified by the equating of the MBE. Once the average proficiency of a group of candidates is determined, scaling will adjust not only for an upswing or downswing in proficiency from past years, but also for any change in the difficulty of written test questions or any change in the harshness of the graders from past years.

In our example of the individual of average proficiency who sits for the bar with a particularly bright candidate pool, this individual's raw written scores will remain lower than they would have been in previous sittings with less able peers. But the equating of the MBE will take into account that this is a particularly bright candidate pool and that the individual in question is in fact of average ability. The individual's written test scores will then be scaled to account for the difference in the candidate pool, and his written test scores will be brought into alignment with his demonstrated level of ability. Scaled essay scores lead to total bar examination scores that eliminate contextual issues and that accurately reflect individual proficiency.

Doesn't this process disadvantage people who do poorly on the MBE?

No. It is important to note that an individual might have one of the best MBE scores and one of the worst essay scores, or vice versa. Scaling written scores to the MBE does not change the rank-ordering of examinees on either test. A person who had the 83rd best MBE score and the 23rd best essay score will still have the 83rd best MBE score and the 23rd best essay score after scaling.

One analogy that might help relates to temperature. Suppose the garages on your street have thermometers that measure their temperatures in Celsius and the houses on your street have thermometers that measure their temperatures in Fahrenheit. The house temperatures range from 66° F to 74° F and the garages range from 19° C to 23° C. Suppose that you have the coldest garage and the warmest house. If the garage thermometers are all changed to the Fahrenheit scale, the temperature readings for all the garages will change, but your garage will still measure the coldest and your house will still

measure the warmest. In Celsius terms, your garage temperature was 19° C; in Fahrenheit terms your garage temperature is 66° F. Either way, it is shown to be the coldest garage. (Note that this example demonstrates only the move to a different scale and the maintenance of identical rank order; it does not demonstrate the adjustments that take place during the equating of exam scores.)

In the example of the individual of average proficiency who sits with an unusually bright candidate pool, scaling the average individual's written test scores to the MBE would not change his rank within the candidate group with which he took the examination. He would still be ranked below where he would have ranked in a less-capable candidate pool. However, his unusually low rank would no longer affect his total bar examination score. The total scores for the entire pool of candidates would reflect what was in fact the case: that it was a particularly bright pool of individuals (i.e., the total scaled scores would be higher than they were for previous administrations).

What does the process of scaling written scores to the MBE entail?

Scaling the written tests can be done either on an individual essay score (or MPT item score), or on the

total written test score. NCBE will scale the scores for individual jurisdictions, if they wish, or will provide the jurisdictions with software to do the scaling themselves. Essentially the process results in generating new scaled essay scores that look like MBE scores. The distribution of written scores will be the same as the distribution of MBE scores—with a very similar average, a very similar minimum and maximum, and a very similar distribution of scores.

The process is described in the Testing Column that appears in the May 2005 BAR EXAMINER (Volume 74, No. 2). Conceptually, the result is similar to listing MBE scores in order from best to worst, and then listing written scores in order from best to worse to generate a rank-ordering of MBE scores and essay scores. The worst essay score assumes the value of the worst MBE score; the second worst is set to the second worst, etc.

Jurisdictions that do not currently scale written test scores to the MBE but are considering doing so should feel free to contact me for more information on how NCBE can help.  scase@ncbex.org

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