Originally appearing in THE BAR EXAMINER*, Vol. 72, No. 2, May 2003. Reprinted with permission of the National Conference of Bar Examiners.*

THE TESTING COLUMN

by Susan M. Case, Ph.D.

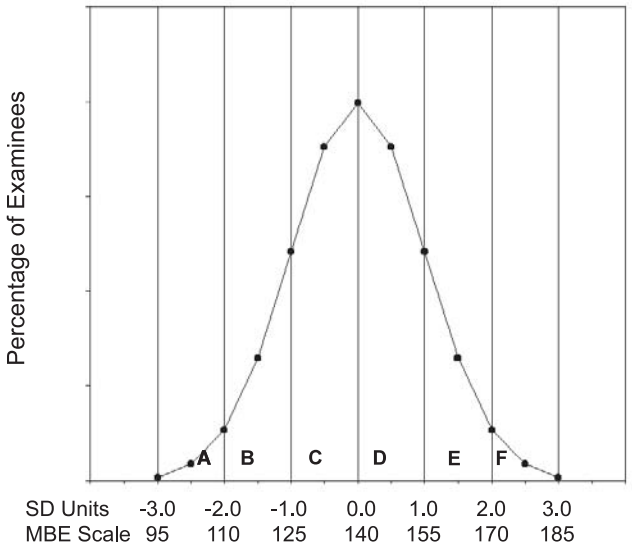
The statistics issue of THE BAR EXAMINER contains a wealth of information of interest to the bar examining community. A series of tables provides bar admissions data from each jurisdiction, including 2002 pass/fail numbers and percentages both overall and broken down by source of legal education and by first-time taker vs. repeater status. It also shows the numbers admitted to the bar by examination, by motion, and by diploma.

This issue presents a new view of the 2002 data for the MBE in the form of a frequency distribution chart. These new graphs were developed to help readers interpret the changes in average scores that we have seen on the MBE over time. This column presents some background information, and then describes more specifically how to read the graph associated with the MBE.

In situations where there are large numbers of examinees (as with the national data on the MBE), the distribution of scores tends to form a bell shape (and is sometimes called a "bell curve"). Each set of scores has a mean (the average score), and a standard deviation or SD that indicates the amount of spread in the scores. The SD for the MBE is about 15 scale points. Scores for most of the examinees fall within 3 SD units below the mean and 3 SD units above the mean. In general terms, this implies that MBE scores for most examinees are between 95 and 185, a range of 90 points (6 x 15).

Figure A below shows a typical bell-shaped curve. The horizontal axis has two labels. The top label ("SD units") shows the points on the scale in terms of standard deviations, ranging from -3.0 SDs to +3.0 SDs. The height indicates the percentage of examinees who scored at a particular level. The highest point of the curve is at "0.0" which indicates the mean of the distribution of scores. Approximately one-third of the examinees have scores between the mean and -1 SD below the mean (the section labeled C);

FIGURE A: Normal "Bell" Curve



approximately one-third of the examinees have scores between the mean and +1 SD above the mean (section D). Approximately 95 percent of the examinees have scores within 2 SDs above and below the mean (sections B, C, D, and E).

The second label for the horizontal axis ("MBE scale") represents a typical set of MBE scores, with a mean of 140 and an SD of 15. The label "140" corresponds to the "0.0" on the SD scale, indicating that 140 is the mean of the MBE scores. The following table shows the translation of MBE scores into SD units.

Scale in SD units	*Scale in MBE units*
-3.0 SDs	(140) - (3 X 15) = 95
-2.0 SDs	(140) - (2 X 15) = 110
-1.0 SD	(140) - (1 X 15) = 125
0.0	(140) - (0 X 15) = 140
+1.0 SD	(140) + (1 X 15) = 155
+2.0 SDs	(140) + (2 X 15) = 170
+3.0 SDs	(140) + (3 X 15) = 185

Figure A shows that the one-third of examinees who are included in Section C had MBE scores between 140 (the MBE mean) and 125 (the MBE mean minus 1 SD); another one-third of the examinees who are included in Section D had scores between 140 and 155 (140 + 15). About 95 percent of the examinees had scores between 110 and 170 (the mean minus 2 SDs to the mean plus 2 SDs); these examinees who fall within plus or minus 2 SDs of the mean are included in the middle four sections of the graph (labeled B, C, D, and E). Most of the remaining 5 percent of examinees are included in the tails of the curve in the sections labeled A and F.

This is a lot of background, included to make a point about how to interpret differences in scores from one examination to another. Generally, a difference of 0.15 SDs or more would be considered meaningful. For the MBE, 0.15 SDs equals 2.25 scale points (0.15 x 15=2.25), so a mean change of slightly over 2 scale points may be considered a meaningful difference.

The normal curve shown in Figure A may also be used to illustrate the impact of a decline in scores on the pass / fail rate. If the pass / fail standard is near the mean, at the peak of the curve, a small change in mean score is likely to have a large impact on the pass / fail rate. This is a result of the fact that, at the peak of the curve, a large proportion of examinees have the same score. On the other hand, if the pass / fail standard is away from the mean in either direction (near either tail of the distribution), a small change in performance will have a much smaller impact on pass / fail rates since few examinees have scores in that range.

FIGURE B: Distribution of MBE Scores for February and July 2002

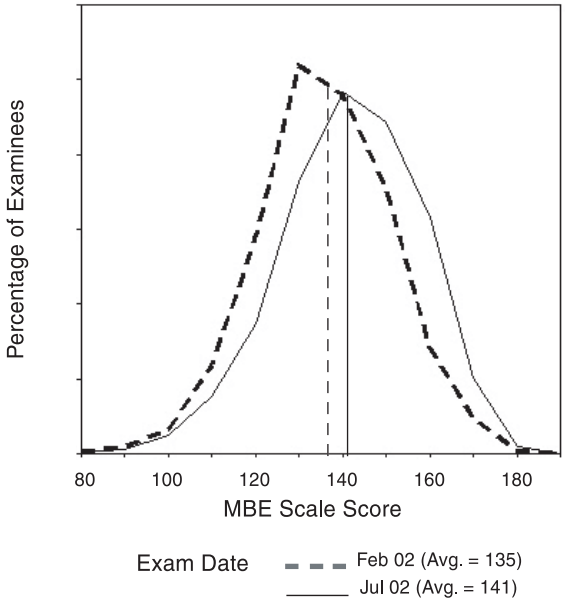


Figure B shows the 2002 MBE results for the February and July administrations. In this graph, the horizontal axis indicates the scaled score, and the vertical axis indicates the percentage of examinees who obtained each scaled score. Vertical lines are drawn at 135 and 141 to show the mean for February (135) and the mean for July (141). Note that the July curve is somewhat to the right of the February curve, illustrating graphically the fact that July scores tend to be higher than February scores. The difference in the average score from February 2002 to July 2002 was 5.9 points on the MBE scale.

The concept of SD units can provide perspective on the magnitude of any differences between performance from one administration to the next. In 2002, the SD for the MBE was about 15.7. The 5.9 point difference between average scale scores in February 2002 and July 2002 is more than one-third of an SD (5.9 / 15.7 = 0.37 SDs). This would be considered a large difference.

FIGURE C: Distribution of MBE Scores for July 2001 and July 2002

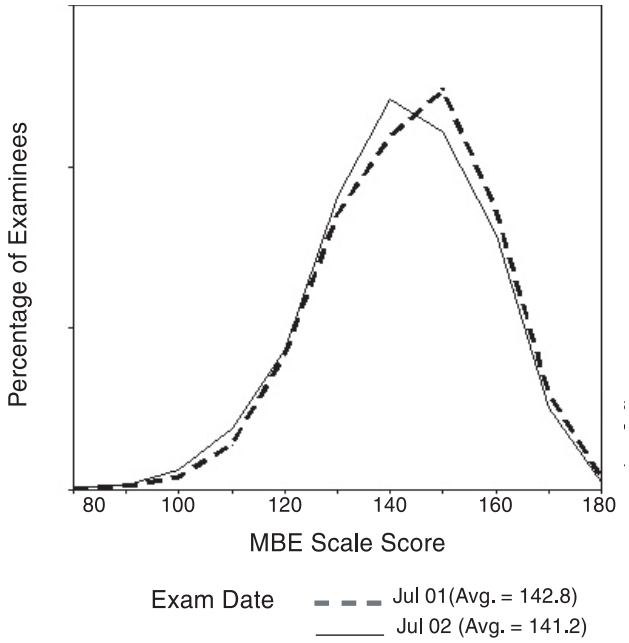


Figure C shows the distribution of MBE scores for the July 2001 and the July 2002 administrations. The difference between the MBE means in July 2002 and July 2001 was 1.6 scale score points, about one-tenth of an SD (1.6 / 15.7 = 0.10 SDs). Note that the two curves are almost identical, with a slight difference in the center of the graph. The differences between scores in Figure C are much smaller than the differences shown in Figure B. In general, the differences between February and July are much greater than between any two July administrations. In fact, the difference between the highest ever July MBE (July 1994) and the July 2002 MBE was four points (145.2 - 141.2 = 4 points); this represents about one-fourth of an SD (4 / 15.7 = 0.25 SDs).

While the concept of SD units can be used to provide perspective on the magnitude of differences in performance, it cannot help in determining why changes have occurred. Detailed analysis of performance would require a database of examinee information that is hypothesized to affect performance. This would include not only the obvious demographic variables such as gender, race, and ethnicity, but also variables such as which law schools examinees attended, their law school GPAs, the number of times they have taken the exam previously, whether English is their primary language, when they graduated from law school, and their LSAT scores. Without background data, we can explore the magnitude of differences across administrations, but we cannot explore the underlying causes of those differences.

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