THE TESTING COLUMN

by Susan M. Case, Ph.D.

As I am writing this column, we are making final plans for the administration of the July examinations, and are just completing analyses of the February Multistate Bar Examination (MBE) results. Some jurisdictions were disappointed with their February results, and asked us to take another look at the data; some jurisdictions provided us with their complete data sets, and asked us to analyze their data in more detail.

First, using national MBE data, we checked to make sure the results were correct. In particular, we checked to make sure that the equating had been done appropriately. A statistical procedure known as equating is used to ensure that an MBE scaled score of 140, for example, represents the same level of proficiency over time. As a result of equating, you can know that an examinee who received a score of 140 on the July 2001 MBE performed at the same level of proficiency as an examinee who received a score of 140 on the February 2002 MBE or on any other MBE administration. Equating allows us to conclude that, if the average score was lower in February, it was because the examinees were less proficient on average, not because the examination was more difficult. Careful re-analyses convinced us that the equating was done correctly and that the results were correct.

Next, we compared the national results of the February 2002 MBE with those of July 2001 MBE. The national mean (average) score for February 2002 was approximately 135, and the national mean for July 2001 was 143, a difference of about 7.5 points. If you prefer to think in terms of standard deviation (SD) units, the difference was about .5 SD (the SD of scores for both February and July was about 15 points). Using either calculation, the difference is big enough for psychometricians to view as significant, and certainly big enough for many to view as important.

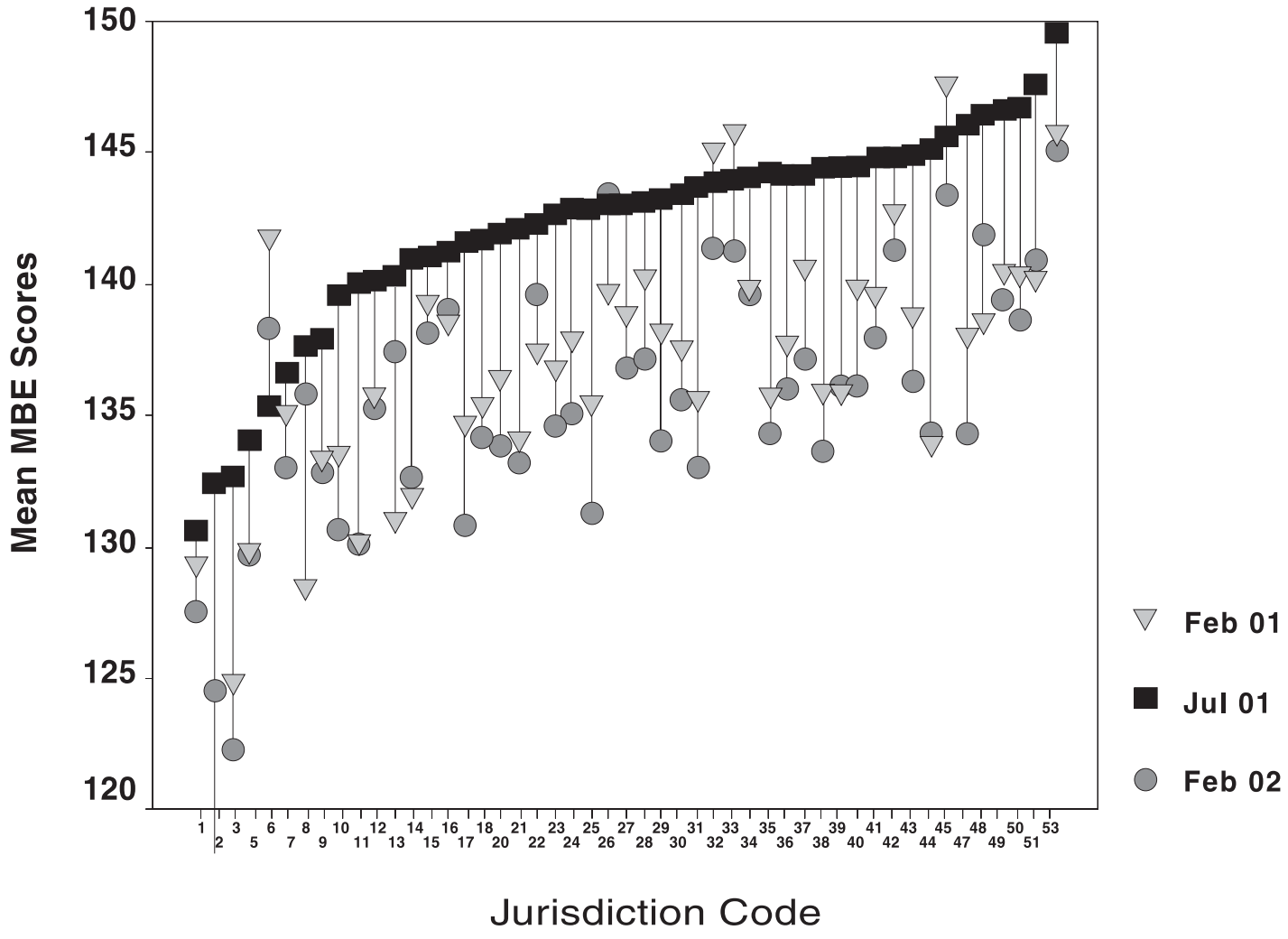
We then reviewed the results over time to see whether this February's result was uncommon. Stepping back a year, the national mean score for February 2001 was approximately 137, about 1 point higher than the mean for February 2002. Looking back even further, the February 2002 results look fairly consistent with the past. Over the past 22 years, the February means have ranged from 135 to 139, while the July means have ranged from 139 to 145. There has been some shifting up and down from one year to the next, but we saw no obvious upward or downward trend.

We also looked at national data on the LSAT to see if those data would help to forecast a trend for the MBE. Some people have noted a decrease in the numbers of applicants to law school, and have

expressed concern that the quality of the applicants has deteriorated. However, the LSAT data, as reported to us, have been remarkably stable over time, and do not reflect a decrease in quality of applicants.

Finally, we examined the results from the February 2001, July 2001, and February 2002 administrations on a jurisdiction-by-jurisdiction basis. As noted above, the mean score nationally in February 2002 was about 1 point lower than in February 2001. However, the results were quite variant across jurisdictions. At the extremes, one jurisdiction had a 7 point drop in mean score between February 2001 and February 2002 while another had a 5 point gain. As expected, most of the jurisdictions had a decrease in mean scores from July 2001 to February 2002 (although one had an increase in the mean); the largest decrease was slightly under 12 points.

The figure below shows mean MBE scores for 47 jurisdictions (eliminating those without February administrations and those with fewer than 10 examinees in July 2001) for those three administration dates. The jurisdictions are ordered by average July



Mean MBE scores for the February 2001 (triangles), July 2001 (squares), and February 2002 (circles) administrations of the MBE for 47 jurisdictions. For example, the first jurisdiction had mean MBE scores of 129 for February 2001, 131 for July 2001, and 128 for February 2002.

2001 MBE score; the jurisdiction codes indicated on the X-axis reflect the rank-order of the jurisdictions, created for the purpose of building this graph. Each horizontal line and trio of symbols represent the data from one jurisdiction. For each trio of symbols, the square represents the jurisdiction's mean MBE score in July 2001, the triangle represents the mean score for February 2001, and the circle represents the mean score for February 2002. The figure illustrates the variability from one jurisdiction to the next. Some jurisdictions experienced quite large differences in mean scores from February to July, while others had much smaller differences. Some had higher scores in February 2001 than in February 2002. A few jurisdictions had higher February scores than July scores.

So, what conclusions can we draw? First, the results appear to be accurate; we have no reason to suspect that there was any problem with the examination itself, the scoring, the equating, or the score reporting. Second, when viewed nationally, the February 2002 results appear to be in line with other February results, although, when viewed on an individual jurisdiction basis, the February results are sometimes quite disparate from one administration to another. This means that care should be taken to avoid over-interpreting both unexpectedly low and unexpectedly high results. The causes of these February swings are likely to be the smaller examinee counts, the higher proportion of repeat examinees, and the higher proportion of examinees who repeated courses or were delayed in graduation. It would be worthwhile to study MBE performance in more detail, but this requires additional data collection and probably some data sharing across institutions.

We are, of course, anxious to receive the July 2002 data. Hopefully, the increase seen from July 2000 to July 2001 will be repeated.

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