

# THE TESTING COLUMN

## WHAT GOES UP OFTEN COMES DOWN: FEBRUARY 2009 PERFORMANCE ON THE MBE

*by Susan M. Case, Ph.D.*

I've received a number of calls regarding the lower scores and lower pass rates on the February 2009 bar examination. As many of you know, MBE scaled scores were at a record high in July 2008, which some hoped was the start of a continued increase in scores. This was not to be. The July 2008 high was followed by a February low, though not the lowest ever.



difficulty into account. As a result, when the average score decreases from July to the next February or from one February to the next February, we can assume that the examinees were less proficient. In this case, examinees in February 2009 were about 2 points less proficient on average than examinees in February 2008, and about 10 points less proficient on average than examinees in July 2008.

Figure 1 shows the trend in average February MBE scaled scores from 1991 to 2009. Based on the total population of MBE takers, mean MBE scaled score changes from one July to the next and from one February to the next have continued to be less than three points. The mean scores for the past 10 February administrations have ranged from 135 to 138; during the same period, July mean scores have ranged from 141 to 146. The July 2008 mean was the highest on record.

Following is an explanation of the February 2009 results in a Q & A format. This article follows directly from the article in the November 2008 BAR EXAMINER'S Testing Column in which the July 2008 results were summarized. Together they are intended to provide an explanation of score increases and of score decreases.

### **1. Does the decreased average score mean the MBE is getting harder?**

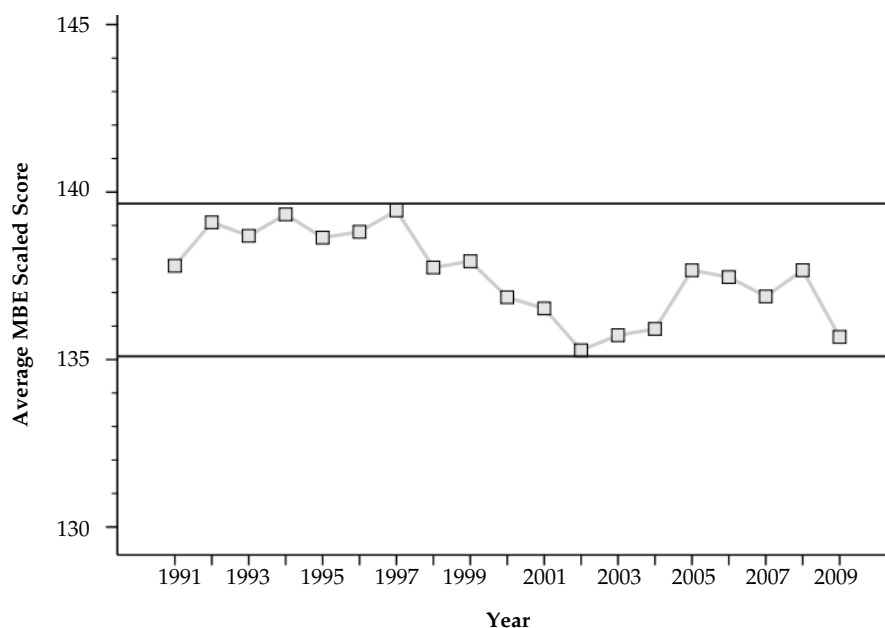
No, the MBE is not getting harder. The MBE scaled score is adjusted to take any differences in question

### **2. What accounts for the decrease in performance?**

Some of the decrease may be attributable to the higher than usual pass rate in July. A second possible explanation is based on background data that we have for about 40 percent of the February 2009 examinees. For this sample, the LSAT scores were lower in February 2009 than in February 2008 both for first-time takers and for repeaters. About 20 percent of the lower MBE scores can be attributed to lower entering scores on the LSAT. However, while we can speculate, we don't actually know what factors resulted in the decrease in proficiency.

Law schools might have some information that would help to shed light on the weaker performance. For example, for their examinee populations, were the LSAT scores lower, were the law school grades lower, or was the proportion of repeaters higher?

**Figure 1: Trend in Average February MBE Scaled Scores**



**3. Did the score decrease apply to all groups evenly?**

We do not yet have this information for a large enough sample to do the analyses. Additional data will be received over the next few months, and we will address this question then. We do know that the decreased MBE scores were seen across the country in about three-fourths of the jurisdictions.

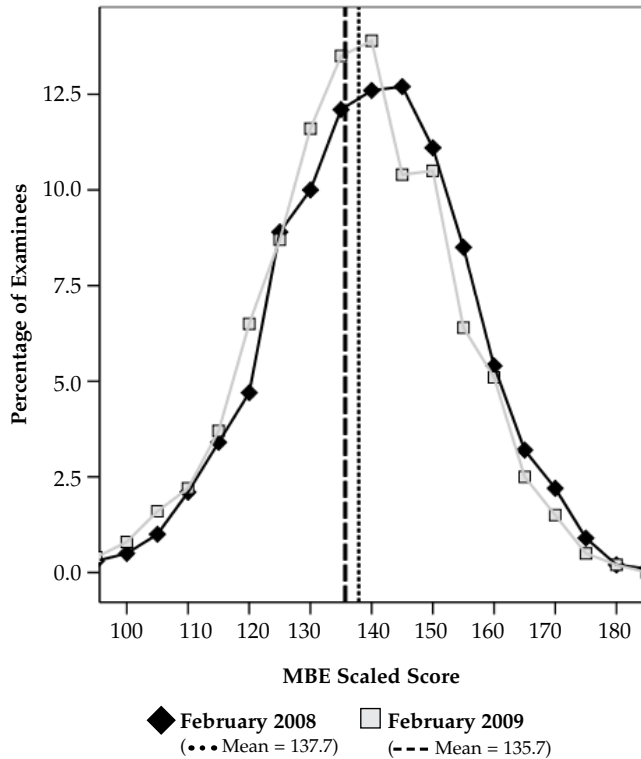
**4. Why was there such a large decrease in the pass rate for some jurisdictions?**

There is always a range of pass rate changes across jurisdictions. In February 2009, approximately three-fourths of the jurisdictions saw a decrease in pass rates from those in February 2008. While mean score decreases for jurisdictions ranged from less than 1 point to over 7 points, decreases in pass rates for these jurisdictions ranged from 1 percentage point to over 16 percentage points. At the same time, a few jurisdictions saw increased scores and pass rates that were similar in magnitude to the decreases described above.

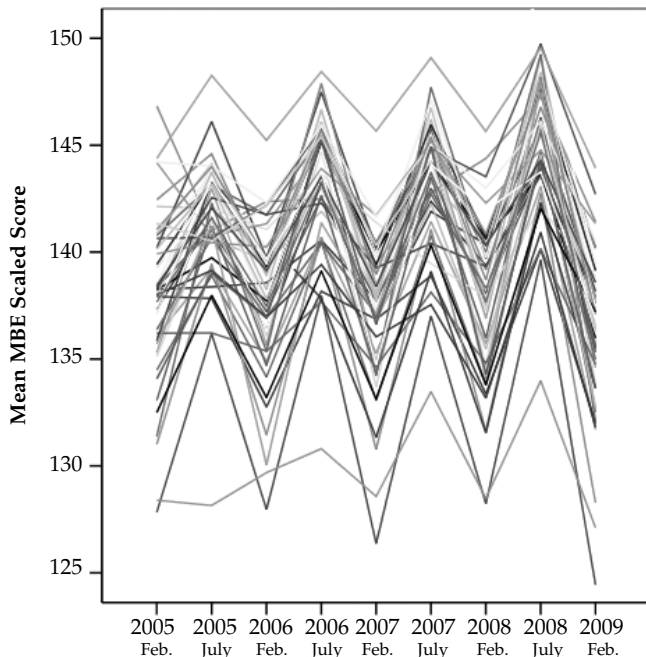
Figure 2 shows the distribution of examinee scores in February 2008 and February 2009. The February 2009 curve is just slightly to the left of the February 2008 curve; it is also slightly more peaked. The curves in Figure 2 highlight how small the difference in performance was. However, although the difference seems small, there are a lot of people represented near the peaks of the curves where most of the pass/fail standards lie. As a result, a small difference in the distribution of scores can create large differences in pass/fail rates.

Note that the reverse phenomenon occurred in July 2008 when the mean score increased and several jurisdictions saw a large increase in the pass rate. Large swings in the pass/fail rates are seen more commonly in small jurisdictions, but large jurisdictions will see large swings in the actual numbers of passers and failers. Figure 3 shows the mean MBE scaled scores for each jurisdiction that tests in both February and July. Each line represents the mean MBE scaled score for a particular jurisdiction. Although specific jurisdictions are not

**Figure 2: Distribution of MBE Scores in February 2008 and February 2009**



**Figure 3: Mean MBE Scaled Scores Across Administrations by Jurisdiction, 2005–2009**




identified in the figure, the variation among jurisdictions and the changes across time are clear.

**5. Could the decreased pass rate be a function of the essay graders being more stringent or the essay questions themselves being harder?**

At this time, almost all jurisdictions scale their essay and performance test scores to the MBE. This procedure is strongly recommended by psychometricians and by NCBE. Those that do scale their written scores to the MBE can be assured that any differences in scores across administrations are not caused by variations in grader stringency, nor are any differences caused by variations in essay question difficulty. Jurisdictions that do not scale their written scores to the MBE cannot determine whether the change in the pass rate is due to a difference in examinee proficiency, a difference in essay question difficulty, or a difference in grader stringency.

Every test administration results in changes to the pass rate; most of these changes are very small, but occasionally mean scores change by several points and the pass rate shifts quite a bit. For jurisdictions that scale their written scores to the MBE, these differences are the result of real changes in examinee performance, not the result of a more or less difficult MBE or a more or less difficult written test. For jurisdictions that scale their written scores to the MBE, these differences cannot be explained by differences in grader stringency or essay question difficulty.

We are left with unanswered questions about why the February 2009 examinees performed less well than usual. As we collect more background information on the examinees, we will revisit these results and try to answer some of the questions that remain.

As always, your comments and questions are welcome, and NCBE is available to assist jurisdictions with these issues. 

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