

THE TESTING COLUMN

DEMYSTIFYING SCALING TO THE MBE: HOW'D YOU DO THAT?

by Susan M. Case, Ph.D.

There seems to be a mystery about how to scale the essay scores to the Multistate Bar Examination (MBE), and the Testing Column of the statistics issue seems to be an appropriate place to run through the process.

Here are the steps in a nutshell:

Step 1. Determine the mean and standard deviation (SD) of MBE scores in your jurisdiction.

Step 2. Determine the mean and SD of the essay scores in the jurisdiction (you can do this for the total essay score, the average essay score, or each essay individually).

Step 3. Rescale the essay scores so that they have a mean and SD that is the same as the MBE for the same group of examinees.

Now for an example. Let's assume you tested fifteen examinees. In Table 1, on the following page, the data for each examinee are shown. The first column shows the examinee ID number. Column 2 shows the total raw essay score that each examinee received. These data were obtained from a mock essay exam where each of the ten essays is graded on a one to six scale, but the calculations apply equally well for any number of essays and any type



of grading scale. The bottom of the column shows the average score (40) and standard deviation of the scores (5.2). These are typically shown in your data output and would not usually be calculated by hand.

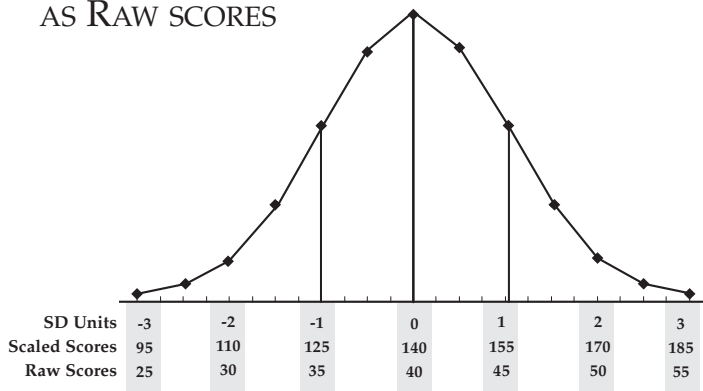
Your calculations begin with column 3. For each examinee, subtract the group mean essay score (40 in this example) from the total essay score that the examinee received; then divide this value by the group SD (5.2 in this example). The result is the total essay score in standard deviation units. For example, the first examinee has a total raw essay score of 29; the mean essay score is 40 and the SD is 5.2; thus, the examinee's score in SD units is -2.1 or 2.1 standard deviations below the mean (see the Testing Column in May 2003 for a discussion of scores in SD units).

Regardless of the score range, scores in standard deviation units typically range from -3 to +3 (zero is the average score). An examinee with a score of -2.1 SDs has a very low score; an examinee with a score of zero SDs has a score at the mean of the group; and an examinee with a score of 2.3 SDs has a high score. Overall, the scores in SD units have a mean of zero and a standard deviation of one; this will always be true if the scores have a normal (bell-curve) distribution (see Figure 1).

TABLE 1. SAMPLE ESSAY DATA SHOWN FOR EACH EXAMINEE

Column 1	Column 2	Column 3	Column 4	Column 5
Examinee ID	Actual Total Raw Essay Score	Total Essay Converted to SD Units	Total Essay Score Scaled to the MBE	Actual MBE Score
1	29	$(29-40) / (5.2) = -2.1$	$(-2.1 \times 15) + 140 = 108.1$	110
2	34	$(34-40) / (5.2) = -1.2$	$(-1.2 \times 15) + 140 = 122.6$	120
3	36	$(36-40) / (5.2) = -.8$	$(-.8 \times 15) + 140 = 128.4$	140
4	38	$(38-40) / (5.2) = -.4$	$(-.4 \times 15) + 140 = 134.2$	140
5	38	$(38-40) / (5.2) = -.4$	$(-.4 \times 15) + 140 = 134.2$	160
6	39	$(39-40) / (5.2) = -.2$	$(-.2 \times 15) + 140 = 137.1$	127
7	39	$(39-40) / (5.2) = -.2$	$(-.2 \times 15) + 140 = 137.1$	149
8	40	$(40-40) / (5.2) = 0$	$(0 \times 15) + 140 = 140.0$	150
9	41	$(41-40) / (5.2) = .2$	$(.2 \times 15) + 140 = 142.9$	140
10	42	$(42-40) / (5.2) = .4$	$(.4 \times 15) + 140 = 145.8$	142
11	42	$(42-40) / (5.2) = .4$	$(.4 \times 15) + 140 = 145.8$	138
12	42	$(42-40) / (5.2) = .4$	$(.4 \times 15) + 140 = 145.8$	130
13	43	$(43-40) / (5.2) = .6$	$(.6 \times 15) + 140 = 148.7$	149
14	45	$(45-40) / (5.2) = 1.0$	$(1.0 \times 15) + 140 = 154.5$	136
15	52	$(52-40) / (5.2) = 2.3$	$(2.3 \times 15) + 140 = 174.8$	170
Mean	40	0	140	140
SD	5.2	1	15	15

FIGURE 1. SAMPLE ESSAY DATA SHOWN IN SD UNITS, AS SCALED SCORES, AND AS RAW SCORES



Column 4 shows the calculations to turn this standard deviation score into a score that is scaled to the MBE. Multiply the examinee’s score in SD units by 15 (the SD of the MBE for this group); add this result to 140 (the mean of the MBE for this group). The mean and SD for the essay scores scaled to the MBE will have the same mean and SD as the MBE scores. Column 5 shows the actual MBE scores achieved by each examinee.

Note that the examinees who performed relatively poorly on the essay have a scaled essay score

that is also relatively low, regardless of their actual performance on the MBE. Examinee 1 performed poorly on both the essay and the MBE. Examinees 3, 4, and 5 performed poorly on the essay but much better on the MBE; note that this comparison can be made easily by comparing the results in columns 4 and 5. Examinee 6 did much better on the essay than on the MBE. Converting the essay scores to the MBE scale did not change the rank-ordering of examinees on the essay scale; it simply made their scores easier to compare. The correlation between the raw essay score and MBE is exactly the same as between the scaled essay score and the MBE.

Scaling the essays to the MBE is an essential step in ensuring that scores have a consistent meaning over time. When essay scores are not scaled to the MBE, they tend to remain about the same: for example, it is common for the average raw July essay score to be similar to the average February score even if the July examinees are known to be more knowledgeable on average than the February examinees. Using raw essay scores rather than scaled essay scores tends to provide an unintended advantage to some examinees and an unintended disadvantage to others.