

REPEATED USE OF MBE ITEMS: ARE SCORES AFFECTED?

by Mary M. Sandifer and Charles S. Kunce

INTRODUCTION

The use of the same test items more than once is a source of concern for ACT, Inc., as well as for the National Conference of Bar Examiners. Of the roughly 61,000 candidates tested per year, around 35% report having taken the Multistate Bar Examination (MBE) previously. Because each form of the MBE contains equating items (questions repeated from previously administered forms), some examinees who repeat the MBE will see some questions again. The issue of whether there is a potential advantage to examinees who are administered these items has policy implications for the Conference and psychometric implications for ACT.

In June and July 1996, ACT investigated the impact of repeat testing on examinees' scores and on item performance. Specifically, ACT designed analyses to determine:

- whether candidates taking the MBE more than once ("repeaters") receive higher scores on test items they have seen before,
- the typical change in score when examinees repeat the MBE, and

- the typical time that elapses between examinees' first and subsequent efforts.

METHOD

The first step was to examine the MBE data files for all jurisdictions for seven consecutive MBE administrations, from February 1993 to February 1996. Examinees who had taken the MBE more than once were identified by matching social security numbers. Throughout this article, *repeater* refers to a candidate whose social security number was found in files of two or more administrations. Matching of social security numbers was performed across all seven administrations in several jurisdictions so that repeaters who moved to another state or repeaters who allowed up to three years to elapse before retesting could still be identified. Records lacking social security numbers were dropped from this analysis. After matching, examinees' records were divided into seven groups corresponding to the number of MBE administrations in which their social security numbers were found; that is, those whose social security numbers appeared once were designated Group 1, twice were Group

ACT, Inc., is under contract to the National Conference for psychometric services related to the development, analysis, and scoring of the Multistate Bar Examination (MBE). The study reported in this article was conducted by ACT, Inc., during the spring and summer of 1996. The National Conference gave permission for this use of its test data, and approved the purposes of the study before it was undertaken.

2, and so on. Numbers of examinees in each of the groups were:

Group 1	97,149
Group 2	15,502
Group 3	3,963
Group 4	1,615
Group 5	689
Group 6	373
Group 7	175

Those who had tested seven times had taken the MBE in a total of twelve jurisdictions.¹ Although this finding was investigated, no particular pattern could be discerned that justified a conclusion other than failed applicants attempting to obtain higher scores, or applicants seeking admission in another, often adjacent, state.

RESULTS

Before February 1993, only California identified examinees by social security number. In February 1996, the last administration included in this study, about half of jurisdictions permitted, but did not require, reporting social security numbers on the answer sheets. Readers must remember that *results reported here are limited to examinees whose social security numbers were contained in test records; that is, those who tested in California, or chose to report their social security numbers.* Despite this limitation, some noteworthy trends appeared and are described in this article.

Trend: Examinees' score increase on repeated test items is negligible.

The advantage that lies in previous exposure to test items proved to be quite modest. Three forms of the MBE were examined for score increases in the equator sets; the February 1996 MBE (Form 296) was double-linked to two

earlier forms, a February form, referred to as Form A, and a July form, referred to as Form B (see Appendix A). Thirty items from each of the linking forms, a total of 60 items, were included in Form 296. Among the examinees who took Form 296, there were non-repeating examinees who had seen neither of the previously administered linking forms, and repeaters who had seen one or both forms. The strategy used to assess the impact of prior exposure to equating items was to isolate three groups of candidates—two groups who had seen one equating set but not the other, and examinees who had seen neither equating set—and compare their performance. The three groups identified were:

Non-repeaters: took neither Form A nor Form B (11,849 examinees)

Repeaters-A: took Form A but not Form B (909 examinees)

Repeaters-B: took Form B but not Form A (892 examinees).

These three groups did not overlap; that is, no one examinee appeared in more than one group.

The relative performance of the three groups on the two linking sets gives information about how memory of items affects test performance. We expected these outcomes:

- *Non-repeaters*, having seen neither set previously, could not be advantaged by memory of the items. Observing this group's performance on the two sets of equating items allowed us to gauge the relative difficulty of these sets. If the scores were similar on each set, the two 30 item sets were considered to be equally difficult.
- *Repeaters-A* would score higher on Form A equators, which they have seen before, than on Form B

Table 1
Mean Raw Scores on Two Sets of Equating Items by Repeater Status

Group	Equating Set A Mean Raw Score	Equating Set B Mean Raw Score	Absolute Difference Between Sets
Non-Repeaters	18.34	18.39	.05
Repeaters-A	17.39	16.97	.42*
Repeaters-B	17.65	18.06	.41*

*statistically significant

equators, if memory of the items gives them an advantage, and if the two equating sets are equally difficult.

- *Repeaters-B* would score higher on Form B equators, which they have seen before, than on Form A equators, if memory of the items gives them an advantage and if the two equator sets are equally difficult.

If there is a benefit of previous exposure to equating items and if the two equator sets are equally difficult, repeat examinees' mean scores on equating items which they *have seen* should be higher than their mean scores on the items which they *have not seen*. Furthermore, the score differences should be statistically significant. These comparisons are summarized in Table 1. Shading indicates which equating set the repeating examinees had taken before.

Non-repeaters' scores on the two item sets showed a negligible difference—.05 points of a possible 30—that was statistically non-significant, indicating that the two equating item sets were equally difficult.

Both groups of repeaters scored slightly higher on the set of equating items which they had taken before. This suggests that a small score increase may occur as a result of prior exposure to test items. However, while the

differences are statistically significant (see Appendix B), the actual magnitude of the average gain (.41 to .42 raw score points of a possible 30) has limited practical significance. Also noteworthy is the fact that both groups of repeaters scored lower on both sets of equators than the non-repeaters.

While the analyses suggest that there is gain associated with prior exposure to the items, the question is, how do such gains affect the MBE testing program? The major concern is whether an applicant who would otherwise fail will pass the bar examination as a result of seeing—and answering correctly—questions he or she has taken before. A failure will become a pass only if an additional point or points causes a score to increase from below the cut point to above it. These data indicate that, on average, a repeat testing candidate must be given 70 to 75 familiar items to be expected to be advantaged by even one score point. The MBE never contains this many repeated items, and the vast majority of repeating candidates see none at all, or, at most, about 30 repeated items.

A repeater has an advantage only if he or she took one or both of the linking forms *and* obtained a score increase that would place the score over the pass mark. The likelihood that a particular repeater will pass instead of fail depends on the interaction of the candidate's ability

with a number of other factors, including the other exams given, each jurisdiction's combining formula and cut scores, as well as whether he or she has previously seen the test items. Bar examiners can examine score distributions in their jurisdictions to see how many passing candidates would have failed if their MBE scores were one or more points lower. They would then be interested in how many examinees in this decision range were repeaters who had taken the particular linking forms for this administration. This group is likely to be quite small or non-existent.

Trend: There were consistent patterns of score increases, time elapsed between tests, and the number of repeats—except for the group that took all seven consecutive tests.

Group 1, for whom no matching ID numbers were found, are assumed to be non-repeaters.² As a reference point for subsequent tables, their mean scores for each administration date are shown in Table 2.

Table 2
Mean Scores for Examinees for Whom Matching Social Security Numbers Were Not Found

	Mean Scale Score	Number
Form 293	143.3	8,153
Form 793	147.8	22,906
Form 294	145.5	5,067
Form 794	150.1	25,700
Form 295	145.5	4,486
Form 795	148.7	23,802
Form 296	141.8	6,975
Total Tested		97,149

In contrast to the non-repeaters whose performance is summarized in Table 2, three distinct patterns were found among repeaters. They were:

- repeaters whose initial scores were higher have larger score increases,
- repeaters with initially low scores retest sooner and more often, and
- repeaters' mean scores increase with each retest, although after the third repeat the gain is slight.

Table 3 is arranged to permit studying scores by test sequence (first, second, etc.) for a particular examinee group, rather than an administration date. An applicant who took the exam twice, Forms 293 and 793 for example, is reported in the same group as another who took the exam twice but on different dates, for example Forms 795 and 296. Notice that the groups' average first scores (or average second scores, or third, etc.) can be compared by reading across rows. An increase in mean score for each group is shown by reading down columns. For example, between the first and second tests, Group 2 (which tested twice) and Group 5 (which tested five times) had increases of 5.7 (133.3 to 139.0) and 2.1 (123.6 to 125.7) respectively. With the exception of Group 7, examinees who repeated the fewest times have higher initial mean scores and their gain upon retesting is greater. This pattern does not apply to Group 7, examinees who tested in the seven consecutive administrations addressed in this study.

Compare the scores of non-repeaters shown in Table 2 with the repeaters scores shown in Table 3. Notice the dramatic difference in mean score between non-repeaters (141.8 to 150.1) and repeaters, both on their first scores (121.6 to 133.3) and their last scores (129.9 to 139.0). In other words, note that the highest mean among the re-

Table 3
Mean Scale Score on the MBE by Number of Times Examinee Has Taken the Examination

	Number of Times MBE Is Taken					
	2	3	4	5	6	7
n	15,502	3,963	1,615	689	373	175
First Test	133.3	127.3	124.2	123.6	121.6	125.6
Second Test	139.0	130.4	126.8	125.7	123.7	124.3
Third Test		135.0	129.1	128.4	126.7	128.7
Fourth Test			132.7	129.2	128.4	129.3
Fifth Test				131.5	129.4	130.1
Sixth Test					130.8	130.3
Seventh Test						129.9

peater group (139.0) is still lower than the lowest mean score among those who did not repeat (141.8).

Trend: The majority of repeaters retest within six months.

Table 3 enables one to read the amount of score change that occurs upon retesting. Also of interest is the typical time that elapses between tests, and how time elapsed is related to score. Percentages of examinees who retest after six months, one year, a year and a half, and so on, were computed for all groups of examinees. Despite the fact that repeat patterns are numerous and can be complicated, a general pattern became apparent (see Appendix C). The majority—over three fourths—of examinees who retest do so within six months, and the lower the initial score, the more likely that the examinee

will retest without delay. Seventy-three percent of Group 2 (repeaters who tested twice) retested within six months, while 94 percent of Group 6 (six-time repeaters) did so. The average initial scale scores of these groups were 133.3 and 121.6 respectively.

In general, the shorter the time that elapsed between tests, the lower the initial scores and the smaller the increase. Readers who are interested in these relationships may compare Table 3 and the information given in Appendix C for a more complete picture of repeat testing patterns. For example, over 90% of Group 6 examinees retest within six months and have a mean score increase of 2.1 scale score points (from 121.6 to 123.7).

Should bar examiners be concerned that repeaters may be advantaged by remembering items that they have

seen six months before? Probably not. Generally, more than six months elapse between repeated uses of equating items. Additionally, repeaters tend to have low initial scores and low gains. For example, on most administrations, a scale score of 124 is in the lower 20 to 25 percent of all examinees. Given the low scores of repeaters and the mean score gains, it appears unlikely that repeaters who retest within six months will pass as a result of having seen the test items.

DISCUSSION

The authors believe that the results of this study should provide a degree of comfort to anyone concerned about whether retesting candidates are unfairly advantaged by previous exposure to MBE test items. One might expect that candidates, after taking the MBE, might try to remember the questions that they were unsure of and study these points of law. However, this research indicates that this apparently doesn't happen. Either candidates do not remember the questions they took, or they do not prepare in their areas of weakness, or some combination of these. One possibility is that because test dates are six months apart, and items may not repeat for several years, candidates have forgotten the questions on the earlier exam when they are preparing for the next one.

The results of this study are reported as information to assist in decision-making, and to prompt additional inquiry. It does not imply specific recommendations for action on the part of the National Conference of Bar Examiners.

ENDNOTES

1. These twelve jurisdictions were: Alabama, California, Colorado, Connecticut, the District of Columbia, Illinois, Kentucky, Maryland, New Hampshire, New Jersey, New York, and Pennsylvania.
2. These scores differ from the scores in the Technical Report because this group of "First Time Examinees" is not identical to the self-reported group in the Technical Report.

MARY SANDIFER, since July 1997, has been Deputy Director of Testing for NCBE. Prior to assuming this position, she was an Assistant Director in the Professional Assessment Services Division of ACT, Inc. In this capacity, she coordinated the development of the MBE and the study that is reported in this article. Dr. Sandifer's background also includes extensive experience with academic testing programs in public education and a Ph.D. in Educational Research with emphasis in measurement and statistics.



CHARLES S. KUNCE has been the Director of the Professional and Occupational Programs Department, Professional Development Division of ACT, Inc. since 1990. In this capacity he oversees the test development activities of a number of high stakes, national licensure and certification examinations, including the MBE, the MEE, and the MPT. Dr. Kunce received his Ph.D. in psychology in 1979 with emphasis in psychometrics and individual assessment.



APPENDIX A EQUATING THE MBE

What is equating?

Equating is an adjustment made to candidates' test scores to neutralize the effects of differences in test form difficulty. This is a procedure that is done in the interest of fairness to candidates. When different forms of a test are developed from one administration to the next, as is done with the MBE, these forms will inevitably fluctuate in their level of difficulty. This happens even if their content specifications are identical. Equating, in effect, corrects for this by giving candidates some score point "credit" if their test form is more difficult.

APPENDIX C

Because many combinations of test dates are possible, only the combinations that account for the majority of repeaters were selected for reporting. For all groups, this proved to be the shortest time that could have elapsed before completing testing, and the shortest time plus six months. For example, the shortest time that could have elapsed between the first and third tests is one year (July-

February-July or February-July-February). One and a half years could have included any three test dates from four consecutive administrations, and so on.

The numbers in the chart are column percentages. That is, for Group 3 which tested three times, 79% of examinees took the second test a half year after they took the first.

Time Elapsed Between Tests		Group Number: number of times examinee has tested					
		2	3	4	5	6	7
Test 1 to Test 2	½ year (2 consecutive admins)	72.9	79.0	84.3	86.9	93.8	100
	1 year	14.8	15.3	11.5	11.8	6.2	
Test 1 to Test 3	1 year (3 consecutive admins)		57.9	70.3	73.7	87.1	100
	1½ year		24.8	18.6	21.6	12.9	
Test 1 to Test 4	1½ year (4 consecutive admins)			55.6	60.5	77.5	100
	2 years			22.6	29.9	22.5	
Test 1 to Test 5	2 years (5 consecutive admins)				43.4	70.2	100
	2½ years				36.7	29.8	
Test 1 to Test 6	2½ years (6 consecutive admins)					57.1	100
	3 years					42.9	