

PR-88-1

ANALYSIS OF THE JULY 1987 EXAM

Stephen P. Klein, Ph.D.
Roger Bolus, Ph.D.

GANSK & ASSOCIATES

February 15, 1988

SUMMARY

The July 1987 exam had two sections: the Multistate Bar Examination (MBE), a 200-item multiple choice test; and a Written Examination composed of six essay questions and two Performance Test (PT) problems. There were 7,481 applicants who completed both sections, 35 percent of whom had taken the GBX at least once before.

MBE raw scores (the number of questions answered correctly) were converted to scale scores that controlled for possible differences in average item difficulty from one administration of the exam to the next. Essay and PT answers were graded on a 40 to 100 point scale. Scores on this scale were assigned in 5-point intervals. Each PT score was then multiplied by 2 so that the maximum possible Written raw score was 1,000 points.

Written raw scores were converted to the same scale of measurement as was used on the MBE. This was done to adjust for possible differences over time in the difficulty of the questions asked and the leniency with which the answers to them are graded. An applicant's total score was computed using the formula below:

$$\text{Total Scale Score} = (.35)(\text{MBE Scale}) + (.65)(\text{Essay Scale})$$

A three phased grading process was used to determine an applicant's pass/fail status. In Phase 1, applicants were: failed if their Total score was less than 1391, passed if their Total was greater than 1465, and placed in reread if their Total was in the 1391 to 1465 range.

In Phase 2, applicants in the reread group had all of their essay and PT answers read a second time. At the end of this phase, applicants were: failed if their Total scores were below 1412, passed if their scores were above 1439, and placed in reappraisal if their scores were in the 1412 - 1439 range. Applicants also were placed in reappraisal if their Total score after either the first or second reading was greater than 1439 (i.e., even if their average was less than 1412).

An applicant in the reappraisal group had all of his or her answers and scores reviewed as a set by a member of the Board of Reappraisers who then made a final pass/fail decision. Thus, an applicant's answers were read once, twice, or three times depending upon how close that applicant came to the pass/fail line of 1440.

The major findings with the 7,481 applicants who had all of their answers read at least once and the group of 1,659 applicants who had them read at least twice were as follows:

- o After the first reading of all answers, 36 percent of the applicants failed and 42 percent passed. The total percent passing after the second reading and the reappraisal process were 48 and 51, respectively.
- o California applicants scored higher than the national average on all of the six MBE subjects. And, the mean California MBE scale score was higher on the July 1987 exam than on most of the 11 previous July California exams.
- o The degree of agreement between readers in their evaluations of the relative quality of the written answers was higher on an essay question than it was on a PT problem. Moreover, because reader assigned PT scores were multiplied by 2, differences in final scores between the first and second reading on a given answer were much larger on a PT than on an essay answer.
- o The reliability of the Written score (.77) was comparable to the reliabilities of written scores on previous exams, and the correlation between MBE and Written scores (.66) also was about the same as it was on these previous tests.
- o The reliability of the Total score was comparable to that of previous July exams.
- o In general, an applicant's total score after the first reading was highly related to both that applicant's total score after two readings and final pass/fail status.
- o The phased grading cutoff scores led to rereading and reappraising the answers of the applicants who were likely to benefit from these additional reviews.
- o Males tended to earn higher scores on the MBE than on the Written section whereas the reverse was true for females. Anglos and Blacks tended to earn slightly higher MBE than Written scores whereas the reverse was true for Hispanic and Asian applicants.

ANALYSIS OF THE JULY 1987 EXAM

TEST SECTIONS

The July 1987 exam had two sections: the Multistate Bar Examination (MBE), a 200-item multiple choice test; and a Written section composed of six essay questions and two Performance Test (PT) problems. The exam was administered over three consecutive days, with two 3-hour sessions per day. Day 1 consisted of essay questions 1 - 3 and PT problem A. Day 2 was devoted to the MBE. Day 3 consisted of essay questions 4 - 6 and PT problem B.

SCORING RULES

MBE raw scores (the number of MBE questions answered correctly) were converted by the American College Testing Program (ACT) to equated ("scaled") scores using the formula below. This procedure adjusts the raw scores for possible differences in average question difficulty from one administration of the MBE to the next.

$$\text{MBE Scale} = (9.2877)(\text{MBE Raw}) + 211.7518$$

Each essay answer was graded in 5-point intervals on a 40 to 100-point scale. The same procedure was used to grade each PT answer. PT scores were then multiplied by 2 so that the maximum possible Written raw score was 1,000 points (6 essays at 100 points each plus 2 PT problems at 200 points each).

Written raw scores were converted to a score distribution that had the same mean and standard deviation as the applicants' MBE scores. This scaling was done using the MBE and Written scores of the random sample of 1,000 applicants who had their Written answers graded first. The formula used to convert written raw scores to scale scores was:

$$\text{Written Scale Score} = 2.7585 (\text{Written Raw Score}) - 410.4324$$

An applicant's Total scale score was a weighted combination of that applicant's MBE and Written scale scores. The formula for computing Total scale scores was:

$$\text{Total Scale Score} = (.35)(\text{MBE Scale}) + (.65)(\text{Written Scale})$$

A three-phased grading process was used to determine an applicant's pass/fail status. In Phase 1, applicants were: failed if their Total score was less than 1391, passed if their Total was greater than 1465, and placed in reread if their Total was in the 1391 to 1465 range.

In Phase 2, applicants in the reread group had all of their essay and PT answers read a second time. The second reader on an answer was a different grader than the first reader and the second reader did not know the score assigned by the first reader.

The scores assigned on the first and second readings were averaged and new Total scale scores were computed using the formulas above. At the end of Phase 2, applicants were: failed if their Total scores were below 1412, passed if their scores were above 1439, and placed in reappraisal (Phase 3) if their scores were in the 1412 - 1439 range. Applicants also were placed in reappraisal if their Total score after either the first or second reading was greater than 1439 (i.e., even if their average was less than 1412).

An applicant in the reappraisal group had all of his or her answers and scores reviewed as a set by a member of the Board of Reappraisers who then made a final pass/fail decision. Thus, an applicant's answers were read once, twice, or three times depending upon how close that applicant was to the pass/fail line of 1440.

SAMPLE

The analyses of the July 1987 exam were conducted with the 7,481 applicants who had an MBE and a Written score. This sample contained 2,582 repeaters and 4,899 applicants who were taking the exam for the first time. The percentage of applicants from ABA approved, California Accredited, and Nonaccredited schools were 66, 18, and 4, respectively.

MULTISTATE BAR EXAMINATION (MBE)

Table 1 shows that California applicants scored higher than the national average on all of the six MBE subtests. California's average total raw score (the average number of questions answered correctly) was 3.83 points higher than the national average. It also was relative high in comparison to most previous July exams in California (see Appendix A).

WRITTEN SECTION

Table 2 presents the means and standard deviations on each question after all readings of all answers. The last column of this table shows the correlation between the scores on a question and the sum of the scores on the other seven questions. The consistent size of these correlations suggests that no question stood out as measuring something quite different than the other questions.

Table 1

NATIONAL AND CALIFORNIA AVERAGE MBE SCORES AND THE
DIFFERENCE BETWEEN THESE AVERAGE SCORES

Test Score	Number of Items	National Mean	CA Mean	Difference
Constitutional Law	30	20.93	21.51	0.68
Contracts	40	25.72	26.20	0.48
Criminal Law	30	18.42	19.26	0.84
Evidence	30	19.91	20.85	0.94
Real Property	30	17.65	17.75	0.10
Torts	40	25.66	26.55	0.89
Total Raw Score	200	128.28	132.11	3.83
NCBE/ACT Scale	200	140.30	143.88	3.58

Table 2

SUMMARY STATISTICAL DATA AFTER ALL READINGS
OF THE ESSAY AND PT ANSWERS (N = 7481)

Question Number	Subject Matter Area	Mean Score	Standard Deviation	Corrected Part-Whole Correlation
1	Evidence	61.97	9.64	.489
2	Civil Procedure	68.13	8.79	.565
3	Remedies	65.21	8.84	.411
4	Trusts	68.19	7.11	.414
5	Criminal Law	66.72	8.25	.514
6	Community Property	67.09	8.22	.490
PT-A	Torts	137.42	14.14	.519
PT-B	Real Property	133.92	14.66	.511

READER AGREEMENT

Applicants who came close to the examination's pass/fail line after one reading of their answers had all of their answers read again by readers who had not graded them previously. Readers were "paired" on an answer to offset possible differences in leniency among them. For instance, if the first reader tended to give relatively low grades (on a substantial number of first read answers), then the second reader was picked from among those who had the opposite tendency. This pairing practice tends to increase observed differences between readers.

Table 3 contrasts the applicants' scores on the first reading with their scores on the second reading. As on past exams, the score assigned to an answer on the first reading tended to be slightly higher than the score assigned to that answer on the second reading.

The last column of Table 3 shows the correlation between the scores on the first and second readings. This coefficient reflects the degree to which the relative standings of the applicants on the first reading were consistent with their standings on the second reading. The higher the coefficient (up to a maximum of 1.00), the stronger the relationship.

The correlations in Table 3 show there was less agreement among readers on PT answers than on essay answers. There was only a .42 correlation between the total written raw score on the first and second readings. However, this correlation as well as those on the individual questions underestimate the degree of agreement between readers because the second readings were restricted to applicants whose total scores after the first reading were in the reread zone.

Table 4 shows each question's distribution of absolute difference scores and average absolute difference scores. The absolute difference is the difference in score assigned to an answer by the two readers who graded it, regardless of the algebraic sign of that difference (e.g., if an answer received scores of 65 and 70 from different readers, then the absolute difference was 5, regardless of which reader graded it first).

Two readers usually differed by about 5 points in the score they assigned to an essay or PT answer. The largest absolute difference was 35 points. This occurred once (on essay question 5). Final absolute differences were about twice as high on PT answers as they were on essay answers because the reader assigned PT scores were multiplied by 2.

RELATIONSHIPS AMONG SECTIONS

Table 5 presents summary statistical data on each section after all readings. The internal consistency of the total score, .891, was based on a formula developed by Gullickson for a weighted composite.

Table 3

AVERAGE ESSAY AND PT SCORES ON THE FIRST AND SECOND READINGS, THE DIFFERENCE BETWEEN THESE AVERAGES, AND THE CORRELATION BETWEEN SCORES ON THE FIRST AND SECOND READING (N = 1659)

Question Number	Mean on First Reading	Mean on Second Reading	Difference in Means	Correlation Between Readings
1	61.4	61.6	-0.2	.58
2	67.5	67.5	0.4	.69
3	64.7	63.9	0.9	.68
4	68.0	67.7	0.3	.64
5	65.8	65.1	0.8	.50
6	66.4	66.7	-0.3	.56
PT-A	137.2	135.9	1.3	.52
PT-B	133.2	132.2	1.0	.34
Total	664.4	660.6	3.8	.42

Table 4

CUMULATIVE PERCENTAGE OF ANSWERS WITH DIFFERENT SIZED ABSOLUTE DIFFERENCE SCORES (N = 1659)

Mean Size of Absolute Difference	Essay Questions						PT	
	1	2	3	4	5	6	A	B
0	33	40	41	42	38	34	42	37
5	73	86	83	88	79	78		
10	91	98	95	97	94	95	84	77
15	97	100	99	100	99	99		
20	99	*	100	*	100	100	98	93
25	100		*	*	*	*		
30				*			99	99
>30					*		100	100
Average Difference	5.3	3.8	4.1	3.7	4.5	4.7	7.7	9.5

* More than 0 but less than 0.5% of the applicants. Differences are for essay answers on a 100-point scale and PT answers on a 200-point scale.

Table 6 contains the correlations among sections and the LSAT. These correlations are biased downwards by the less than perfect reliability of the measures; e.g., if all the measures were perfectly reliable, then the correlation between MBE and Written scores would be .81.

MBE and Written scores correlated about as highly with each other on the July 1987 exam as they did on previous exams. About 75 percent of the applicants had the same pass/fail status on the MBE as they had on the Written section (where a scale score of 1440 on a section was considered passing).

SUBGROUP ANALYSES

Table 7 shows that on the average, women scored 30 points higher on the Written section than on the MBE whereas there was a 28 point average differential in the opposite direction for male applicants. Anglos and Blacks scored higher on the MBE than on the Written section whereas the reverse was true for Asian and Hispanic applicants. The true size of the difference between sections for Black applicants is hidden somewhat by the relatively smaller percentage of males in this group.

PHASED GRADING

As noted previously, a three-phased grading process was used to focus reader time on the applicants who came close to the pass/fail line. Table 8 presents the number and percentage of applicants in each of the exam's six pass/fail categories. These data indicate that 3,780 (50.5 percent) passed.

Table 9 shows the number of Phase 2 applicants who passed and failed after all readings relative to their Phase 1 scores. This table illustrates the strong, but far from perfect relationship between Phase 1 scores and final pass/fail status. These data also indicate that the width of the Phase 2 reread band (1391 - 1465) was about right in that almost all applicants with high scores on the first reading passed and almost all with low initial scores failed. For instance, only 5 of the 117 applicants with first read scores below 1395 passed the exam.

Similarly, Table 10 shows that the higher an applicant's score at the end of Phase 2, the greater the likelihood that applicant passed as a result of reappraisal. The lower limit of the reappraisal band (1412) also appears to be appropriate in that only 1 of the 58 applicants with Phase 2 Total scores below 1415 passed as a result of reappraisal.

Table 5

SUMMARY STATISTICAL DATA AFTER ALL READINGS (N = 7481)

Test Statistic	MBE Scale	Written Raw Score	Total Scale
Average Score	1438.8	668.6	1435.8
Standard Deviation	153.6	51.1	133.5
Internal Consistency	.866*	.771	.891

* Computed by ACT using national data.

Table 6

CORRELATIONS AMONG SECTIONS AND LSAT
AFTER ALL READINGS (N = 7481)

	LSAT*	MBE	Written
MBE	.55		
Written	.53	.66	
Total	.59	.86	.95

*LSAT (Law School Admissions Test) scores
were available for 6515 applicants.

Table 7

MEAN SCALE SCORES WITHIN RACIAL/ETHNIC AND SEX GROUPS AND THE NUMBER OF APPLICANTS AND PERCENTAGE OF MALES WITHIN EACH GROUP*

Test	Racial/Ethnic Group				Sex	
	Anglo	Asian	Black	Hispanic	Female	Male
MBE	1459	1383	1331	1368	1421	1450
Written	1451	1396	1328	1378	1451	1422
Total	1454	1392	1330	1374	1440	1432
Number of Applicants	5895	462	481	513	3001	4443
% Male	60	63	49	63	0	100

* There were 93 applicants who did not fall in the four largest racial/ethnic groups and 37 applicants who did not provide data about their racial/ethnic or sex group.

Table 8

NUMBER AND PERCENTAGE OF APPLICANTS WHO PASSED AND FAILED IN EACH PHASE OF THE MULTIPHASED GRADING PROCESS

Pass/Fail Category	Number	Percent
Fail - Phase 1	2709	36.2
Fail - Phase 2	566	7.6
Fail - Phase 3	<u>426</u>	<u>5.7</u>
Total Fail	3701	49.5
Pass - Phase 1	3113	41.6
Pass - Phase 2	484	6.5
Pass - Phase 3	<u>183</u>	<u>2.4</u>
Total Pass	3780	50.5
Phase 1 = 1st reading	5822	77.8
Phase 2 = Reread	1050	14.0
Phase 3 = Reappraisal	609	8.1

Table 9

NUMBER OF REREAD APPLICANTS WHO PASSED AND FAILED RELATIVE TO THEIR TOTAL SCORES AFTER THE FIRST READING (N = 1659)

Score After 1st Reading	Number of Applicants			Percent Passing
	Fail	Pass	Total	
1450 - 1465	49	277	326	85
1435 - 1449	131	206	337	61
1420 - 1434	190	115	305	38
1405 - 1419	297	44	341	13
1390 - 1404	325	25	350	7
Total	992	667	1659	40

Table 10

NUMBER OF REAPPRAISED APPLICANTS WHO PASSED AND FAILED RELATIVE TO THEIR TOTAL SCORES AFTER REREAD (N = 586)

Total Score After Reread	Number of Applicants			Percent Passing
	Fail	Pass	Total	
1435 - 1239	31	82	113	73
1430 - 1434	45	65	110	59
1425 - 1429	74	18	92	20
1420 - 1424	98	11	109	10
1415 - 1419	98	6	104	6
<1415	57	1	38	2
Total	403	183	586	31

Appendix A: HISTORICAL EXAM RESULTS

Year	FEBRUARY			JULY		
	Mean MBE	Number of Applicants	Percent Passing	Mean MBE	Number of Applicants	Percent Passing
1976	1367	3088	38	1453	6709	60
1977	1387	3399	44	1430	7191	55
1978	**	**	**	1447	6835	55
1979	1390	4166	45	1440	7152	55
1980	1367	3758	34	1417	7379	49
1981	1380	3837	33	1420	7080	50
1982	1373	4033	31	1427	7038	49
1983	1360	4200	28	1437	7277	50
1984	1370	3899	30	1407	7201	42
1985	1403	4661	33	1417	7622	46
1986	1397	4689	28	1437	7780	45
1987	1406	4682	43	1439	7481	51

Data were not available for the February 1978 exam. The February 1987 exam was the first one on which written scores were scaled to the MBE.