

AN ANALYSIS OF THE JULY 1983 CALIFORNIA BAR EXAMINATION

Prepared by:
Stephen P. Klein, Ph.D.

March 17, 1984

PURPOSE

The July 1983 version of the California bar exam had three sections: an Essay test, the Multistate Bar Examination, and two Performance Tests. The analyses described in this report were conducted to answer the following questions about this exam:

- o Was the scoring of the exam's essay and Performance Test written answers as reliable as the scoring of essay answers on prior exams; i.e., did different readers tend to assign the same score to a given answer?
- o Was the total score across all three sections sufficiently reliable for making pass/fail decisions about individual applicants?
- o Was any section more difficult than any other section?
- o Was the total exam as difficult as previous California exams?
- o Were the exam's phased grading decision rules as accurate as the rules used on past exams?

SCORING PROCEDURES

The theoretical possible score range on each section was 0 to 600 points. This portion of the report discusses the general procedures that were used to administer and score each section, including the scaling of the multiple choice parts of the exam.

Essay Test

The Essay portion was given in two, 3-hour sessions on the first day. There were three questions per session. Some questions dealt with a single content area whereas others covered more than one area. There was a separate team of 12 Essay answer readers per question. The scores on a question were assigned in 5-point intervals up to a maximum of 100 points. Applicants were assigned a score of at least 40 if they appeared to make a serious attempt to answer a question.

Multistate Bar Examination (MBE)

The MBE was given in two, 3-hour sessions on the second day. The MBE contains 200 multiple choice items (questions) from six content areas, with four choices per question. It is given as part of the bar exam in almost every jurisdiction in the country. MBE raw scores (the number of items answered correctly) are scaled by the Educational Testing Service to control for the possibility that the items asked on one exam may, on the average, be more or less difficult than those asked on another exam.

The scaling of MBE scores is accomplished by including in the group of 200 items, a subset that had been used previously but are still secure. These repeated items are called "equators" in order to distinguish them from items that had not been asked previously. Standard statistical formulas are used to compare the performance of past and current applicants on the equators. The results of this comparison are then used to determine how total raw scores on the July 1983 exam should be adjusted so that they could be compared fairly with scale scores earned on prior MBEs. The formula for converting July 1983 MBE total raw scores to scale scores is:

$$\text{California MBE Scale Score} = 2.6460(\text{total raw score}) + 83.9841$$

All applicants benefited from the scaling. For instance, an applicant who answered 130 items correctly (65 percent of those asked) received an MBE scale score of 428 (which is 71 percent of the theoretical maximum score of 600) rather than the 390 that applicant would have earned if the scaling had not been done.

Performance Test (PT)

Two PTs were given on the third day. Applicants were allowed 3.5 hours per PT. Each PT had a "library" of fictitious cases and statutes in one booklet and a "file" of memoranda, interview notes, documents, and transcripts in another booklet. The morning PT dealt with a torts case and the afternoon PT with a civil procedures case.

Each PT had two parts, a written test and a multiple choice test. The theoretical maximum score on the written and multiple choice portions of a PT were 200 and 100 points, respectively.

The written portion of the morning PT asked the applicant to use the library and file to prepare a memo that presented an objective analysis of the strengths and weaknesses of their client's case as well as a brief description of the additional information that should be gathered. In the afternoon, applicants were instructed to use the library and file for the second problem to write a persuasive argument in favor of their client's case in the form of a Memorandum of Points and Authorities.

Each written portion was graded in 5-point intervals up to a maximum of 100 points. Applicants were assigned a score of at least 40 if they made a serious attempt to answer to this portion of the PT. An applicant's raw score on the written portion was multiplied by 2.0 so that the theoretical maximum score on a written answer was 200 points. There was a separate team of 16 readers for each PT.

The multiple choice portion of a PT contained 15 questions. A question typically asked applicants to evaluate various cases in the library with respect to how much they would support a legal principle that was relevant to their client's case.

Raw scores on the multiple choice part of each PT were scaled so that the mean and standard deviation of the scale scores would be one sixth as large as the corresponding statistics on the MBE. The formulas for converting raw PT multiple choice scores to scale scores appear below. Note that these formulas do not adjust a given applicant's PT score in terms of that applicant's MBE score and they are based on parameters that were derived from all the applicants who took the MBE and PT.

$$\text{PT-AM Scale Score} = (4.5088 \times \text{AM raw score}) + 20.9988$$

$$\text{PT-PM Scale Score} = (3.4692 \times \text{PM raw score}) + 32.4314$$

The decision was made to adjust scores on the PT's multiple choice sections to a score distribution that was based on the distribution of MBE scale scores because:

- (1) There was no way of knowing in advance of the exam whether the PT multiple choice items were unduly easy or difficult and with only a total of 30 items, a few of them could have a large adverse affect on the passing rate.
- (2) The distribution of MBE scale scores provided an objective and independent basis for determining the difficulty of the PT items because its scale scores were already adjusted for possible variations in average question difficulty from one exam to the next.
- (3) The procedures for scoring the multiple choice sections could be announced in advance of the administration of the exam and thereby avoid any suspicion that the Committee adversely influenced the percent passing.
- (4) A previous study (Klein, 1981a) indicated that applicants who had high scores on the machine scorable portion of the PT's prototype also tended to have high MBE scores (the correlation was .38). On the July 1983 exam, there was a .55 correlation between the MBE and the average PT multiple choice score.
- (5) Scaling to the MBE rather than to the Essay was expected to increase the percent passing the whole exam because studies of past exams (Klein, 1983) indicated that about 21 percent more applicants passed the MBE than passed the Essay and the mean written scores on prototype versions of the PT were no higher than mean essay scores (Klein, 1981a).

PASS/FAIL DECISION RULES

The basic pass/fail rule on the July 1983 exam was that an applicant had to have a total score (Essay + MBE + PT) of 1260 or higher in order to pass. Previous research (Klein, 1980 and 1982) had indicated, however, that it was possible to identify with an extremely high degree of accuracy a large portion of the passing applicants without grading all of their essay answers. The resources saved by reading only some of the answers could then be devoted to additional readings of the answers of applicants whose initial scores placed them near the pass/fail line. The decision rules in this multiphased grading process were:

- 1) If the sum of an applicant's MBE score, score on two randomly selected essay questions, and scale scores on the multiple choice sections of the PT was greater than 739, then the applicant passed. If this sum was not greater than 739, the applicant went to Phase 2. About 67 percent the applicants who took all three sections of the exam and passed, passed in this phase (see Appendix A).
- 2) If the sum of an applicant's MBE score, score on the first reading of all six essay questions, and scores on both PTs was less than 1213, then the applicant failed. If this sum was greater than 1278, the applicant passed. If it was between 1213 and 1278, the applicant went to Phase 3.
- 3) Applicants in Phase 3 had all of their essay answers and written parts of their PTs graded again by readers who had not seen them before. The average of the first and second readings was then used to compute a Phase 3 total score. If this total was less than 1236, the applicant failed; if it was greater than 1259, the applicant passed; and if it was between 1236 and 1259, the applicant went to Phase 4.
- 4) An applicant in Phase 4 had all of his/her answers and scores reviewed by a member of California's Board of Reappraisers. As a result of this review, the overall passing rate was increased by about 3 percent (see Appendix A).

Prior to the July 1983 exam, applicants were allowed to pass in two ways: by having a total score (MBE + Essay) over the pass/fail line and/or by passing the MBE on one administration of the exam and the Essay on another administration. Contrary to the Committee's expectations when this policy was first adopted, it had the effect of lowering rather than increasing the percent passing. Moreover, it was impractical to extend the rule to the July 1983 exam because it would require a "trifurcation" policy. However, because some July 1983 examinees were repeaters, they were allowed to grandfather a previous pass.

The grandfathering rules were: A repeater who earned a passing score on the MBE between February 1982 and February 1983 was deemed to have passed the exam if that applicant earned a score of 840 or higher on the balance of the test. An applicant who passed the Essay during this same period only needed to earn an MBE score of 420 or higher in order to pass. Applicants did not jeopardize a previous pass by retaking a section they had passed previously. Applicants also could transfer an MBE score from another jurisdiction if that score was earned before the July 1983 exam.

POPULATION AND ANALYSIS SAMPLE CHARACTERISTICS

There were 7,277 July 1983 applicants who had scores on all three sections of the exam. A random sample of 865 of these applicants had all of their Essay and PT written answers read twice as part of the regular grading process regardless of their scores in Phases 1 or 2. Applicants in this sample did not have their actual pass/fail status affected by the additional readings required for this research.

Table 1 presents data on the characteristics of the applicants in the analysis sample and the population. These data indicate that the sample is very representative of the population of examinees who took all three sections of the exam. For instance, average MBE scores differed by less than one point and the percent passing by only 1.2 percent. Except as otherwise noted, the remainder of this report is based on analyses with the sample of 865 applicants.

READER RELIABILITY

Because all 865 applicants in the analysis sample had their Essay and PT answers independently read twice, it was possible to assess whether the PT readers were any more or less consistent with each other than were Essay readers. This analysis focused on the degree to which two readers were likely to assign the same score to a given answer and the degree to which they were likely to rank order the answers in the same way (correlation).

Table 2 presents the cumulative percentage distribution of absolute difference scores between the first and second reading of an answer. For instance, if the first reader assigned a score of 70 to an answer, the absolute difference would be 5 regardless of whether the second reader assigned it a score of 65 or 75. These data indicate that more than 90 percent of the answers had a difference of less than 15 points between readers.

The next to the last row of Table 2 indicates that the score assigned by one reader to an answer usually differed by about 5 points from the score assigned by another reader to that same answer. The last row contains the correlation between the first and second reading of an answer. These data show that the Essay and PT readers were equally reliable.

The means and correlations in Table 2 are consistent with those obtained in interreader studies of past exams (Klein, 1981c and 1982). For instance, the average correlation between the first and second readings of the answers on an essay question was .77 on both the February 1982 and the July 1983 exams.

Finally, the average score on the first reading of an Essay or PT written answer did not differ substantially from the average score on the second reading of that answer. Table 3 shows that the small differences that did occur were balanced out by summing across questions. The total scores on the first and second reading of the Essay section were 403 and 401, respectively. The corresponding averages on the PT written sections (after multiplying each score by 2.0) were 270.4 and 270.7.

Table 1

SAMPLE AND POPULATION CHARACTERISTICS

Variable	Sample	Population	Difference
Number of Applicants	865	7,277	---
Average Essay score[a]	402.1	400.8	1.3
Average MBE score	431.1	430.6	0.5
Average PT MC score[b]	71.8	71.9	0.1
Percent Passing	51.2	50.0	1.2
Percent Male	65.9	64.1	1.8
Percent Anglo	80.3	80.7	0.4
Percent First Timers	75.8	72.9	2.9
Percent ABA Graduates	63.5	64.8	1.3

[a] The population's average essay score was estimated using the essay scores assigned in Phase 1. It was not possible to use this procedure to estimate the average score on the written portion of the PT in the population because all applicants did not have at least one of their PT written answers graded.

[b] This is the average of the two scaled scores on the multiple choice portion of the PT.

Table 2

CUMULATIVE PERCENTAGE OF ANSWERS WITH DIFFERENT SIZED DIFFERENCE SCORES AND CORRELATIONS BETWEEN FIRST AND SECOND READINGS OF THE ESSAY ANSWERS

Size of Absolute Difference	Essay Questions						PT		Average Cumul. %	
	1	2	3	4	5	6	AM	PM	Essay	PT
0	36	29	39	37	36	45	28	37	36.8	32.6
5	79	71	85	78	79	89	69	82	80.2	75.2
10	95	91	97	93	95	98	90	95	94.9	92.2
15	99	98	99	98	98	100	97	99	98.8	97.8
20	100	100	100	99	100	*	99	100	99.8	99.5
>20	*	*	*	100	*		100	*	100.0	100.0
Average difference	4.6	5.6	4.0	4.7	4.6	3.4	5.9	4.4	4.5	5.2
Correlation	.79	.66	.78	.74	.80	.84	.76	.83	.77	.80

*Less than one percent of the answers had a difference of this size.

Table 3

AVERAGE SCORES ON FIRST AND SECOND READINGS OF ESSAY AND PT WRITTEN ANSWERS

Section	Question	First	Second	Difference
Essay	1	70.01	69.06	0.95
Essay	2	65.98	65.43	0.53
Essay	3	68.40	67.84	-0.56
Essay	4	65.24	64.60	0.64
Essay	5	67.25	67.69	-0.44
Essay	6	66.20	66.43	-0.23
Essay	Average	67.18	66.84	0.34
PT	1	66.22	67.69	-1.47
PT	2	69.00	67.64	1.36
PT	Average	67.61	67.67	-0.06

It may be concluded from these data that there was no appreciable difference between the Essay and the PT in scorer reliability (consistency of grading). And, the reader agreement rates on both the Essay and the PT were just as high as those obtained in studies of prior California exams.

INTERNAL CONSISTENCY

Internal consistency reliability refers to the degree to which applicants who do well on some questions in a test also tend to do well on the other questions in that test. Reliability is an important test characteristic because it can indicate the degree to which chance may have affected applicant scores. Examples of chance factors are guessing on multiple choice items and studying a case just before the exam that is especially relevant to one of the exam's questions. A test or a combination of tests should have a reliability of about .90 if scores on it are used for making pass/fail decisions about individuals.

The internal consistency of the July 1983 MBE was .885 in a random sample of applicants from across the country. This is typical of the reliability coefficients obtained on previous versions of the MBE.

The internal consistency of the Essay section in the sample of 865 applicants was .774 when all six of their Essay answers were read twice. This degree of consistency is almost as high as that obtained on previous exams when each of nine essay answers was read only once. For example, estimates of the internal consistency coefficients of the total essay score on the July 1980 and 1982 exams were .81 and .79, respectively.

The reliability of the total PT score was .698 when both of its written portions were read twice and combined with the multiple choice scores. This coefficient was estimated from the .525 correlation between the morning and afternoon total scores. The reliability of the written portion of the PT was .671 (based on the .505 correlation between the average AM and average PM written scores).

Item analyses conducted by the Educational Testing Service with a random sample of 950 applicants from the population of applicants who took the PT indicated that the internal consistency of its AM and PM multiple choice sections was .28 and .57, respectively. The lower reliability of the AM section was due to two items that were answered correctly by over 95 percent of the applicants and two items that had much lower than average correlations with other 13 items in the test.

The reliability of the total score when all the Essay and PT written answers were read twice was .919 (as determined by standard formulas for computing the reliability of a linear composite). Thus, total exam scores exceeded the minimum standards for reliability.

RELATIONSHIPS AMONG SECTIONS

Table 4 presents the correlations among the various sections and PT subsections of the exam. These data indicate that MBE and Essay scores correlated slightly higher with each other than they did with PT scores and

that the relative standings of the applicants on the written portion of the PT was by far the major determiner of their relative standings on the PT (the part-whole correlation between the PT written score and PT total was .96 whereas the part-whole correlation between the PT multiple choice and PT total was only .66). The right hand column of Table 4 shows that all three sections correlated to about the same degree with total bar scores.

RELATIVE DIFFICULTY OF SECTIONS

Table 5 shows the average score and standard deviation on each section. These data indicate that the MBE was easier than the PT which in turn was easier than the Essay. For instance, the percentage of applicants with scores of 420 or higher on the MBE, PT, and Essay were: 61.3, 46.0, and 30.6, respectively. Figure 1 shows the distribution of MBE, Essay, and PT scores. The horizontal lines in each bar represent, from top to bottom, the 90th, 75th, 25th, and 10th percentile points. The asterisk (*) shows the median (50th percentile). For example, about 50 percent of the applicants had MBE scores at or below 433 whereas 50 percent had Essay scores at or below 402.

The difference in relative difficulty between the MBE and Essay (30.7 percent with "passing" scores) was larger than the usual 20 to 22 percent differences between these two sections on previous exams. Because the PT's difficulty level fell about midway between the MBE and the Essay, its inclusion in the exam did not appreciably affect the overall percent passing.

As noted previously, scaling the MBE raw scores had the effect of increasing an applicant's score on this section. Adjusting the PT multiple choice scores in terms of the distribution of MBE scale scores generally had the opposite effect because, on the average, the PT multiple choice items were easier than those asked on the MBE. These differences in average item difficulty were eliminated by adjusting PT multiple choice scores in terms of the distribution of MBE scale scores. If no scaling had been done on either the MBE or the PT, then the percent passing the exam would have been substantially lower because: (1) the MBE carried three times more weight in determining an applicant's total score than did the sum of the two PT multiple choice sections and (2) the average per item size of the correction on the MBE was larger than it was on the PT multiple choice sections.

RELATIVE DIFFICULTY OF THE EXAM

About 46.9 percent of the applicants in the analysis sample had total scores of 1260 or higher before reappraisal (Phase 4). The reappraisal process and in the case of one applicant, the transferring of an MBE score from another jurisdiction, increased the total percent passing to 51.2 percent. None of the applicants in the analysis sample passed as a result of the grandfathering of the bifurcation rule.

The total percent passing in the analysis sample and the population was higher than it had been on the two prior July exams.

Table 4

CORRELATIONS AMONG SECTIONS, PT SUBSECTIONS,
AND TOTAL BAR EXAMINATION SCORES

	Performance Test					Total GBX
	MBE	Essay	Multiple choice	Written	Total	
MBE	---	.73	.55	.57	.64	.91
Essay		---	.46	.61	.65	.90
PT-MC			---	.42	.66	.63
PT-Written				---	.96	.81
PT-Total						.86

Total GBX = MBE + Essay + Total PT

Table 5

SUMMARY TEST STATISTICS IN SAMPLE OF 865
APPLICANTS WHO HAD ALL OF THEIR ESSAY
AND PT WRITTEN ANSWERS READ TWICE

	Average	Standard Deviation	Internal Consistency
MBE	431.07	48.75	.885
PT	414.14	42.73	.689
Essay	402.08	35.81	.774
Total	1247.29	112.68	.919

Internal consistency coefficient for PT and
Essay are based on two readings per answer.

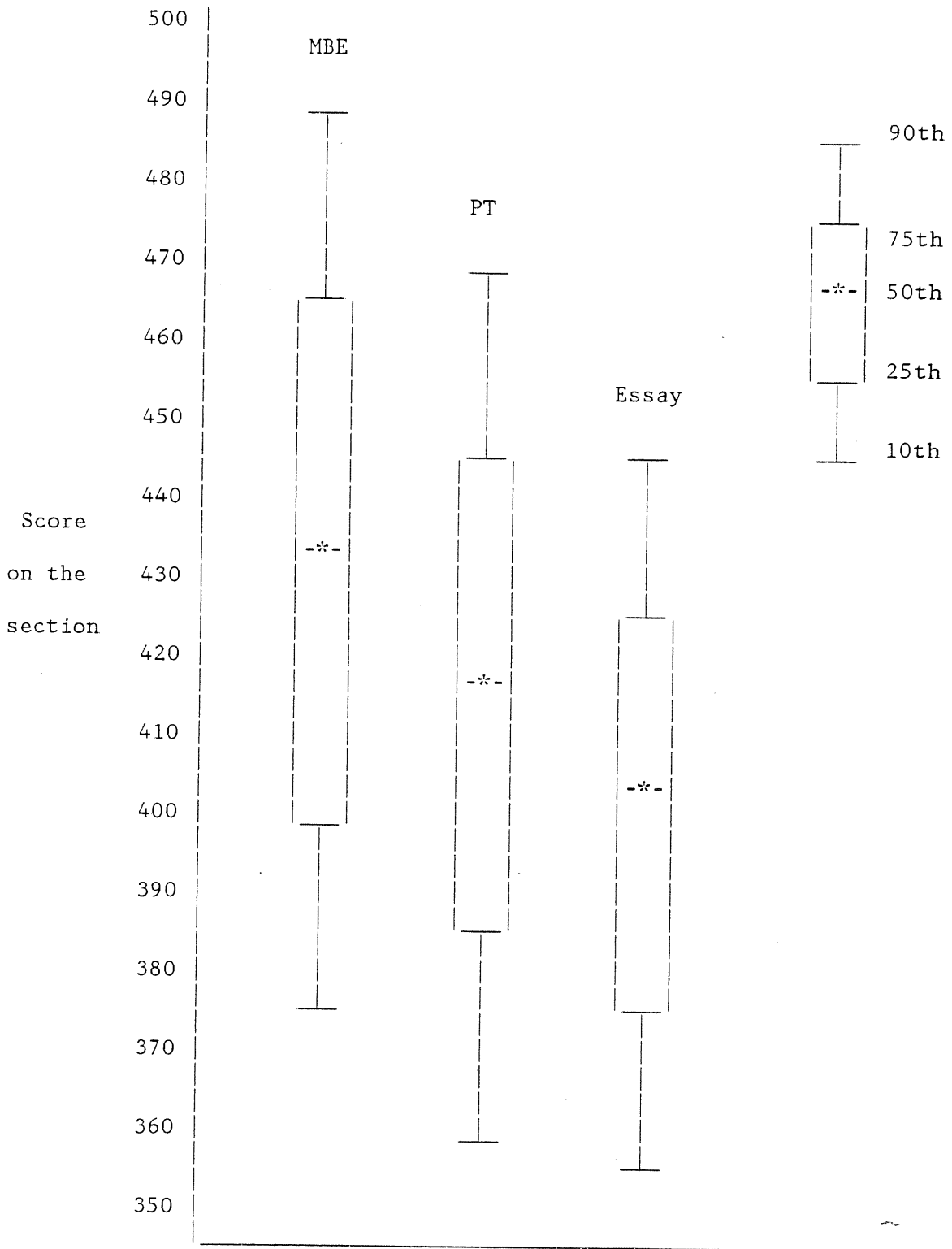


Figure 1. The 10th, 25th, 50th, 75th, and 90th Percentile Points on Each Section.

A difficulty index was computed for the July 1983 exam in order to assess whether its grading standards were more or less lenient than those employed on prior exams. The procedures used to compute this index were slightly different from those used in the past (Klein, 1983) because the July 1983 exam contained three rather than just two sections. And, the index had to be computed on just the analysis sample rather than the full population who took all three sections because as a result of the phased grading process, all the applicants in the population did not have at least one of their two PT written answers read. However, considering how representative the analysis sample was of the population (see Table 1), it is very unlikely that these factors biased the computed value of the index.

The steps involved in computing the difficulty index were: scale the Essay and PT scores to a distribution that had the same mean and standard deviation as the MBE scale scores, compute the sum of each applicant's three scale scores, divide this sum by 9 (to put it on the same 200 point scale that had been used in the past), tabulate the distribution of these average scores, and then find the score in this distribution that would result in 51.2 percent of the applicants passing.

The results of this analysis yielded a difficulty index of 143.6 for the July 1983 exam. The average index on the seven prior July exams was 142.3 (Klein, 1983). It may be inferred from these data (and the 14.4 standard deviation of the average scale scores) that the July 1983 exam was not significantly more difficult than prior California bar exams.

ACCURACY OF PHASED GRADING RULES

The bar exam only has to classify applicants into two categories, pass and fail. There is no need to know by how much an applicant passed or failed. This situation led California to adopt a multiphased grading system that concentrates reader resources on the applicants whose pass/fail status is truly in doubt rather than spreading these resources equally among all applicants taking the exam. For instance, all applicants do not have all of their essay answers read at least once whereas others may have all their answers read as many as three times.

The decision to read all of an applicant's Essay and PT written answers once is based on that applicant's scores on two randomly selected essay answers, the MBE, and the multiple choice sections of the PT. The decision to read them all again is based on the sum of all the scores after the first reading. The decision to read them all a third time is based on the sum of the scores after the second reading.

Data on the sample of 865 applicants who had all their Essay and PT written answers read twice (regardless of their initial scores) provided an opportunity to check the appropriateness of the cutoff scores for each phase of the grading process. The general question addressed by the analyses described below was: Would the phased grading system make the same pass/fail decisions as would have been made if all the applicants had gone through all of the first three phases of it?

Phase 1

The Phase 1 decision rule was: pass all applicants who have a total score of 740 or higher on two randomly selected essay answers, the MBE, and the scaled multiple choice sections of the PT. About 34 percent of the 865 applicants in the analysis sample had Phase 1 scores of 740 or higher.

The appropriateness of the Phase 1 cutoff score was tested by counting the number of applicants who would have been passed in Phase 1, but who nevertheless did not pass after all their answers had been read at least twice. Because there were 6 essay questions, there were 15 possible combinations of two of them. For example, a Phase 1 score might involve essay questions 1 and 2, 1 and 3, 2 and 3, etc.

A test of the efficacy of the Phase 1 cutoff score with each of the 15 possible pairs of essay questions indicated that on the average, it would classify 1.5 applicants out of 865 as having passed when in fact they did not pass after at least two readings of all of their answers. This is an error rate of .0017 (which is well below the .0050 rate that the Committee of Bar Examiners considered to be tolerable). The Phase 1 rule precludes failing an applicant in this phase. Thus, no errors were made in this direction.

In a related analysis, it was discovered that 1,877 applicants (25.8 percent of the 7,277 who had scores on all three sections) had MBE scores of 465 or higher and that only 16 of the applicants in this group failed the exam. Thus, California may want to consider revising the Phase 1 rule so that it focuses on just the MBE portion of the exam.

The use of the PT multiple choice score in Phase 1 contributed to the accuracy of the Phase 1 cutoff score. This conclusion is supported by the .66 correlation between the PT multiple choice score and the total PT score and by the .63 correlation between the PT multiple choice and the total score on the examination.

Phase 2

A Phase 2 error would occur if an applicant's pass/fail status at the end of this phase would have been reversed in Phase 3 as a result of the second reading of all of that applicant's essay and PT written answers. Thus, two types of Phase 2 errors could be made: passing an applicant who should have failed and failing an applicant who should have passed. The likelihood of making each type of error was investigated by counting the number of applicants who had a Phase 2 score:

- o greater than 1278, but whose total scores were less than 1260 after a second reading of all of their answers
- o less than 1213, but whose total scores were greater than 1259 after a second reading of all of their answers

The results of this analysis indicated that there was no one in either group. All the applicants with Phase 2 scores over 1276 had Phase 3 scores over 1260. Thus, the upper bound of the Phase 2 cutoff (1278) was set at about the right spot.

The lowest Phase 2 score that when averaged with the scores from the second reading resulted in the applicant having a Phase 3 score of 1260 or higher was a 1242. The sample of 865 applicants contained 81 examinees with Phase 2 scores between 1213 and 1242. Only 3 of these applicants eventually passed...they earned a Phase 3 score that was high enough to put them into reappraisal and they eventually passed as a result of Phase 4. The Phase 2 scores of these three applicants were 1223, 1234, and 1240.

In summary, the upper bound of the Phase 2 band identified all the applicants who would have passed if all of their answers were read twice. The lower bound did not not exclude anyone from Phase 3 (and thereby from Phase 4 as well) who would have earned a passing score if all of their answers were read twice. Moreover, the lower bound erred on the side of sending more applicants to Phase 3 than were likely to benefit from the double reading. For instance, if the lower bound was raised from 1213 to 1221, it would eliminate the replicate readings of all the answers written by about 3 percent of the total applicant pool. It also would reduce the number of applicants going on to reappraisal.

Phase 3

Phase 3 placed applicants into three categories: pass, reappraise, and fail. The accuracy of the pass/reappraise decision was tested by seeing if there was any applicant with a Phase 3 score over 1260 who did not pass the exam. This analysis indicated that there was not a single applicant that the complete phased grading procedure classified as a fail who would have passed if it was required for all applicants to have all of their essay and PT written answers read twice.

The accuracy of the fail/reappraise decision was tested by determining how many applicants with Phase 3 scores just above 1235 eventually passed. If several applicants with scores in this zone passed, it would suggest that the reappraisal cutoff was too high; i.e., more applicants should have gone to Phase 4. Conversely, if there was no one in this zone, it would suggest the cutoff was too low; i.e., fewer applicants should have gone to Phase 4.

Table 6 shows the number of applicants with Phase 3 scores between 1235 and 1259 relative to their final pass/fail decision. These data indicate that five applicants with Phase 3 scores between 1235.0 and 1244.5 passed. However, the table also shows that only one of these applicants passed because of reappraisal. The other four passed in Phase 1. Given the reappraisal decisions on all 26 applicants in the 1235.0 to 1244.5 zone, these four applicants are probably best categorized as Phase 1 errors rather than as indicators that the bottom of the Phase 3 band was set too high. The one applicant in the 1235-1244.5 zone who did pass as a result of Phase 4 had a Phase 3 score of 1242.

The foregoing findings suggest that all the applicants who might have benefited from having their answers read twice did indeed have them read twice and the Phase 3 cutoff score was too liberal in the sense that it sent more applicants to reappraisal than was really necessary.

Table 6

NUMBER OF APPLICANTS WITH PHASE 3 SCORES BETWEEN 1235.0 AND 1259.5
WHO PASSED AND FAILED THE EXAM IN PHASES 1 AND 4

Phase	Decision	1235.0	1240.0	1245.0	1250.0	1255.0	Total
		to 1239.5	to 1244.5	to 1249.5	to 1254.5	to 1259.5	
1	Pass	1	3	1	4	3	12
4	Pass	0	1	2	6	9	18
4	Fail	11	10	10	7	2	40
Total		12	14	13	17	14	70

Eight of the 39 applicants with Phase 3 scores between 135 and 145 passed the exam, however, 5 of these 8 applicants passed as a result of Phase 1. In other words, only 3 of the 34 applicants in Phase 4 who had Phase 3 scores below 1250 passed as a result of reappraisal. All but 2 of the 11 applicants in Phase 4 who had Phase 3 scores above 1254.5 passed as a result of reappraisal. Thus, the Phase 3 score is a very good, but not a perfect predictor of reappraisal decisions.

SUMMARY AND CONCLUSIONS

The July 1983 California bar exam had three sections: a six question Essay test, the Multistate Bar Examination (MBE), and two Performance Tests (PTs). The theoretical score range on each section was 0 to 600 points.

MBE raw scores (the number of questions answered correctly) were scaled in order to adjust for possible differences in average question difficulty from one exam to the next. This scaling benefited all applicants; i.e., raw scores were consistently lower than scaled scores.

Raw scores on the multiple choice portions of the PT were scaled to the MBE because among other reasons, there was no way of determining in advance of the exam whether the PT's items would be unduly easy or difficult. Scaling to the MBE rather than the Essay or the PT's written score also increased the percent passing the examination.

There were 7,277 applicants who had a score on all three sections of the exam. A random sample of 865 of these applicants had all of their Essay and PT written answers read twice during the regular grading process. Analyses of their data indicated the following:

- o Readers agreed with each other on the score that should be assigned to a PT written answer to the same degree that they agreed with each other on the score that should be assigned to an essay answer. The average difference between two readers on an answer was about 5 points.
- o The degree of interreader agreement on the July 1983 exam was the same as that obtained on prior California exams.
- o The total score on the exam (MBE + Essay + PT) was sufficiently reliable for making pass/fail decisions.
- o The MBE was the easiest part of the test and the Essay was the hardest. For example, the percentage of applicants with scores of 420 or higher on the MBE, PT, and Essay were: 61.3, 46.0, and 30.6, respectively. The tendency for the MBE to be easier than the Essay is consistent with the findings on past exams, but the size of the difference was larger than usual.
- o The net effect of scaling the raw MBE and PT multiple choice scores was to increase the percent passing the exam because all applicants benefited from the MBE's scaling and the MBE carried three times as much weight as the sum of the two PT multiple choice sections in determining an applicant's total score.
- o The percent passing the July 1983 exam was slightly higher than the percent passing the two previous July exams. This increase was probably due to a slight increase in the legal abilities of those taking the exam because the July 1983 test was not more or less difficult than previous exams.

- o The .0017 misclassification rate in Phase 1 was below the .0050 rate set by the Committee. In other words, fewer applicants were misclassified as passes in Phase 1 than was anticipated. Equally accurate Phase 1 decisions could be made for 26 percent of all the applicants solely on the basis of their MBE scores.
- o The parameters for Phases 2 and 3 did not fail any applicant who would have passed if all of his/her essay and PT written answers were read twice. However, the Phase 2 and 3 cutoff scores did lead to more replicate readings than was minimally necessary to make accurate phased grading decisions. Thus, the phased system erred slightly on the side of being too liberal.

REFERENCES

- Klein, S. A comparison of the effectiveness of a single versus multiphased grading system. Report prepared for the Committee of Bar Examiners of the State Bar of California, 1980.
- Klein, S. Testing research skills on the California Bar Examination. Report prepared for the Committee of Bar Examiners of the State Bar of California and the National Conference of Bar Examiners, 1981a.
- Klein, S. An analysis of possible variations in pass/fail standards on the California bar examination. Report prepared for the Committee of Bar Examiners of the State Bar of California, 1981b.
- Klein, S. The effect of time limits, item sequence, and question format on applicant performance on the California bar exam. Report prepared for the Committee of Bar Examiners of the State Bar of California and the National Conference of Bar Examiners, 1981c.
- Klein, S. Analysis of the February 1982 bar examination. Report prepared for the Committee of Bar Examiners of the State Bar of California, 1982.
- Klein, S. Sources of variation in passing rates on California bar exams. Report prepared for the Committee of Bar Examiners of the State Bar of California, 1983.

APPENDIX A

NUMBER OF APPLICANTS IN EACH PASS/FAIL CATEGORY RELATIVE TO THEIR SCHOOL TYPE IN THE POPULATION OF APPLICANTS WHO TOOK ALL THREE SECTIONS OF THE JULY 1983 EXAMINATION

Pass/fail Category	ABA Approved	California Accredited	Not Accredited	All Others	Total
Fail - Phase 2	879	762	197	358	2,196
Fail - Phase 3	597	273	49	104	1,023
Fail - Phase 4	251	63	11	33	358
Fail - but passed Essay in Phase 4	44	10	2	5	61
Pass - Phase 1	2,027	155	20	227	2,429
Pass - Phase 2	532	73	9	61	675
Pass - Phase 3	229	44	4	14	291
Pass - Phase 4	131	31	6	14	182
Pass because of prior MBE pass	27	19	3	13	62
Total	4,717	1,430	301	829	7,277

None of the applicants who took all three sections passed as a result of combining a previous essay pass with a current MBE pass.