

AN ANALYSIS OF THE PERFORMANCE TEST ON THE JULY 1983 CALIFORNIA BAR EXAMINATION

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

May 14, 1984

BACKGROUND

Several studies conducted in conjunction with the July 1980 version of California's General Bar Examination suggested it would be possible to assess some important legal practice skills that were not measured by the regular Essay and multiple choice portions of the exam. Applicants who had clinical legal experience scored slightly higher on these "performance" tests than applicants who did not have such experience and most applicants felt the new tests were more realistic and better measures of their ability to practice law than the traditional parts of the exam. These findings and other factors, such as the increased emphasis on clinical skills in law school, led California's Committee of Bar Examiners to include two of the new measures on the July 1983 exam.

PURPOSE

The analyses described in this report were designed to provide answers to the questions below about the Performance Test (PT) portions of the July 1983 exam:

- o Did the inclusion of the PT significantly affect who passed the exam as a result of the PT measuring skills and abilities that were not already tested by the essay and multiple choice sections of the exam?
- o Did the PT significantly affect the percentage of applicants passing the exam? In other words, was the PT more or less difficult than the essay and multiple choice sections?
- o Was the effect of including the PT on the exam the same for ABA and non-ABA graduates, for different sex and racial/ethnic groups, and for repeaters and first timers?
- o What were the measurement characteristics of the PT? For instance, were readers as consistent with each other in assigning PT scores as they were in assigning essay scores?
- o Did applicants like the idea of including the PT in the exam?
- o After controlling for general legal ability, did applicants with prior legal practice experience score higher on the PT than applicants who did not have such experience?

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. This score was then multiplied by two. There was a separate team of 16 readers for each PT, most of whom had graded answers on a prototype of the PT that was given in conjunction with the July 1980 exam.

Pass/fail Policy

An applicant could theoretically earn 1800 points on the whole exam; 600 on the essay (6 questions with 100 points per question), 600 on the MBE, and 600 on the PT sections (2 PT's worth 300 points each, 100 for a multiple choice part and 200 for a written part). California's basic pass/fail rule was that an applicant had to earn a total score of 1260 or higher in order to pass the exam. A description of how California operationalized this policy by means of a multiphased grading and reappraisal process is presented in another report (Klein, 1984).

Sample

There were 7,277 July 1983 applicants who took all three parts of the exam. A random sample of 865 of these applicants had all of their Essay and PT written answers read twice by reinserting their answers into the regular grading process. The second reader was different than the first one and did not know the score assigned by the first one. Applicants in this sample did not have their actual pass/fail status affected by the additional readings required for this study.

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 parts of the exam. For instance, average MBE scores differed by less than one point and the percent passing by less than one percent. Except as otherwise noted, the remainder of this report is based on analyses with this sample of 865 applicants.

PRELIMINARY ANALYSES

Preliminary analyses were conducted with the sample of 865 applicants to assess certain measurement characteristics of the MBE, Essay, and PT parts of the exam that might affect their relationships with each other and their impact on pass/fail decisions. These analyses investigated: reader reliability (i.e., whether the PT readers were more or less consistent in their grading practices than Essay readers), internal consistency (e.g., whether applicants who had high scores on the morning PT also tended to have high scores on the afternoon PT); and the relative difficulty of each part of the exam.

The next portion of this report describes: the three sections of the July 1983 exam, the general procedures used to score each section, and the characteristics of the sample used in the analyses discussed in the remaining portions of this report. A technical analysis of the 1983 exam is presented in another report (Klein, 1984).

PROCEDURES

Tests and Scoring

The July 1983 version of California's bar exam had three parts: an Essay test, the Multistate Bar Examination (MBE), and two PTs. Six hundred points were assigned to each part.

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 effort to answer a question.

The MBE was given in two 3-hour sessions on the second day. It contained 200 multiple choice items with four choices per item. The test covered six content areas, such as torts and contracts. The MBE 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 to control for possible variations in average item difficulty across administrations of the exam. California multiplies the values in the scaling formula by 3.0 so that the theoretical maximum California MBE score is 600 points. The conversion formula for the July 1983 exam was:

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

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 (AM) PT dealt with a torts case and the afternoon (PM) PT with a civil procedures case. Each PT had a 15-item multiple choice test and a written test.

A PT multiple choice 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 each PT multiple choice test were converted to a score distribution whose mean and standard deviation were one sixth as large as the California MBE's mean and standard deviation. This procedure gave a PT multiple choice score one sixth the weight as the MBE in determining an applicant's total score. The formulas for making this conversion were:

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

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

Table 1
SAMPLE AND POPULATION CHARACTERISTICS

Variable	Sample	Population	Difference
Number of applicants	865	7,277	---
Average Essay score	402.1	400.8	1.3
Average MBE score	431.1	430.6	0.5
Average PT MC score*	71.8	71.9	0.1
Percent passing before reappraisal	46.9	47.5	0.6
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

* This is the average scale score on the multiple choice portion of the PT. It was not possible 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.

Table 2

AVERAGE ABSOLUTE DIFFERENCE SCORES AND CORRELATIONS BETWEEN READERS, AVERAGE SCORES ON FIRST AND SECOND READINGS, AND THE DIFFERENCE IN AVERAGE SCORES BETWEEN READINGS

Section	Question	Between Readings		Average Scores		
		Average	Corre- lation	First	Second	Difference
		Absolute Difference				
Essay	1	4.6	.79	70.01	69.06	0.95
Essay	2	5.6	.66	65.98	65.43	0.53
Essay	3	4.0	.78	68.40	67.84	0.56
Essay	4	4.7	.74	65.24	64.60	0.64
Essay	5	4.6	.80	67.25	67.69	-0.44
Essay	6	3.4	.84	66.20	66.43	-0.23
Essay	Average	4.5	.77	67.18	66.84	0.34
PT	1	5.9	.76	66.22	67.69	-1.47
PT	2	4.4	.83	69.00	67.64	1.36
PT	Average	5.2	.80	67.61	67.67	-0.06

Average absolute difference = difference between two readings of an answer without regard to the algebraic sign of that difference

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).

The third column in Table 2 shows the average absolute difference in 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 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 next column contains the correlation between the first and second reading of an answer.

The remaining columns in Table 2 show that 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. The small differences that did occur were balanced out by summing across questions: the total score on the first and second reading of the Essay section were 403 and 401, respectively; and, the corresponding averages on the PT written sections were 270 and 271.

Taken together, the data in Table 2 indicate that the Essay and PT readers were equally consistent in their grading practices and that the scores assigned on the second reading were not systematically higher or lower than those assigned on the first reading.

Internal Consistency

Internal consistency reliability refers to the degree to which applicants who do well on some of the 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 affect the degree to which scores on one test correlate with scores on another test. For instance, if a test has a low internal consistency as a result of many applicants guessing which answer was correct, then scores on this test would not correlate well with scores on some other measure. A test (or a combination of tests) should have a reliability of about .90 if scores on it are used for making important 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. There was a .65 correlation between the average total Essay score in the morning and afternoon sessions. The average correlation between any two Essay questions was .37.

The reliability of the total PT score (when both of its written portions were read twice) was .689 (as 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).

The correlation between the AM and PM multiple choice sections in the analysis sample was .30. It was .28 in the population of 7,277 applicants. Item analyses of the PT multiple choice sections with a random sample of 950 cases from the population indicated that the internal consistency of the AM and PM 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 (which effectively reduced the length of the test to 13 items) and to two items that had much lower than average correlations with other items in the test.

The reliability of the total exam score was estimated to be .919 when all the Essay and PT written answers were read twice and when each section of the exam was weighted equally in determining an applicant's total score.

Relative Difficulty

Table 3 shows the average score and standard deviation on each section of the exam. These data indicate the MBE was much 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. 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.

RESULTS

Relationships Among PT Scores

Table 4 shows the correlation between PT multiple choice and written scores by problem and across problems. These data indicate that the AM multiple choice scores correlated just as highly with the AM written scores ($r = .24$) as they did with the PM written scores ($r = .23$). This finding suggests that the relationship between PT multiple choice and written scores is not affected by the unique characteristics of each problem. However, the reliability of the AM PT score was quite low (.28).

The reliability of the multiple choice score was much higher on the PM problem than on the AM problem. And, the PM multiple choice scores correlated higher with the PM written scores ($r = .43$) than they correlated with the AM written scores. This finding suggests that there may be a unique effect due to problems and that this effect becomes more detectable as the reliability of the multiple choice scores increases.

The correlations in the last column of Table 4 indicate the written scores carried substantially more weight than the multiple choice scores in determining an applicant's relative standing on the PT. In fact, an

Table 3

SUMMARY TEST STATISTICS IN SAMPLE OF 865

	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.

Table 4

CORRELATIONS AMONG PT SUBSECTIONS

	Multiple Choice			Written			PT Total
	AM	PM	Total	AM	PM	Total	
Multiple Choice							
Morning	---	.30	.79	.24	.23	.27	.47
Afternoon		---	.82	.28	.43	.41	.59
Total			---	.79	.82	.42	.66
Written							
Morning				---	.50	.96	.93
Afternoon					---	.96	.92
Total						---	.96

applicant's PT written score was almost a perfect predictor of that applicant's total PT score. These findings along with the only slightly greater reliability of the total PT score over the written score suggest that it may not be necessary to continue to have a PT multiple choice section (although it did prove useful for California's phased grading process).

Relationships Among Sections

The most important issue addressed by this research was whether including the PT on the exam affected who passed the exam. This issue was investigated by assigning a pass/fail status to an applicant solely on the basis of that applicant's combined MBE and Essay score and then computing the percentage of applicants whose status changed as a result basing the pass/fail decision on the combination of the MBE, Essay, and PT scores.

Because the large differences in difficulty among the sections would have biased results; scores on each section were scaled so that the Essay and PT had the same mean and standard deviation as the MBE (prior to its being multiplied by 3.0). A pass/fail line of 145.2 was used for the analysis because it resulted in about the same percent passing a section as actually passed the exam prior to California's reappraisal process.

Table 5 presents the results of this analysis. These data indicate that basing the pass/fail decision on the MBE+Essay+PT score versus just the MBE+Essay score changed the pass/fail status of 8.4% of the applicants. Specifically, including the PT resulted in 4.3% passing who would not have passed if the exam had just contained the MBE and Essay whereas 4.1% failed because the exam contained the PT. These changes are about the same as would have occurred if the pass/fail decision had been based on just the Essay+PT score or the MBE+PT score rather than the combination of all three scores. This pattern suggests that the difference in the abilities measured by the MBE and Essay is comparable in size to the difference between the MBE and the PT, and between the Essay and the PT.

The same conclusion may be drawn from a cross-tabulation of pass/fail decisions when these decisions are based on the PT versus the MBE or the Essay (see Table 6). For instance, about 77 percent of the applicants had the same pass/fail status on the MBE as they had on the Essay (13 percent passed the MBE only and 10 percent passed the Essay only). This agreement rate is only slightly higher than that obtained between the MBE and PT (72 percent), and between the Essay and the PT (73 percent).

The fact that the PT had a lower internal consistency than either of the other two parts of the exam raised the question of whether this characteristic may have contributed to its having a lower agreement rate than that observed between the MBE and the Essay. This possibility was explored by computing the correlation among the three parts of the exam both before and after the scores on these parts were corrected for their lack of perfect reliability.

Table 7 shows that there would have been a substantial increase in the correlation between sections if their scores were perfectly reliable, but that (1) the correlations with the PT never went over .90 and (2) the MBE and Essay still correlated higher with each other than they did with the

Table 5

PERCENTAGE OF APPLICANTS WHOSE PASS/FAIL STATUS WAS AND WAS NOT CLASSIFIED IN THE SAME WAY BY DIFFERENT COMBINATIONS OF SCORES WHEN THE SECTIONS WERE MADE EQUALLY DIFFICULT TO PASS

	MBE+Essay vs MBE+Essay+PT	MBE+PT vs MBE+Essay+PT	Essay+PT vs MBE+Essay+PT
Both scores agree pass	43.0	42.9	43.5
Both scores agree fail	48.7	48.8	49.6
Total Agreements	91.7	91.7	93.1
Pass by total score only	4.3	4.4	3.8
Pass by subscore only	4.1	3.9	3.1
Total Disagreements	8.4	8.3	6.9

Column percentages may not sum to 100.0% due to rounding.

Table 6

PERCENTAGE OF APPLICANTS WHOSE PASS/FAIL STATUS WAS AND WAS NOT CLASSIFIED IN THE SAME WAY BY DIFFERENT SECTIONS OF THE GBX WHEN THE SECTIONS WERE MADE EQUALLY DIFFICULT TO PASS

First test	Second test	Pass both	Fail both	Total agreed	Pass first only	Fail first only	Total disagreed
MBE	Essay	36.9	40.1	77.0	12.8	10.2	23.0
MBE	PT	34.3	37.6	71.9	15.4	12.7	28.1
Essay	PT	33.6	39.5	73.1	13.4	13.4	26.8

Row percentages may not sum to 100% due to rounding.

PT. This means that if the three parts the exam were equally difficult, including the PT would have a slightly larger impact on who passed than would adding either another day of essay or MBE type testing. However, the substantial increase in the correlations among the measures when they are corrected for reliability suggests that the 8.4% difference in pass/fail decisions that occurred when the PT was added to the MBE and Essay slightly overestimates the PT's unique contribution to those decisions.

Table 7

OBSERVED AND CORRECTED CORRELATIONS AMONG SECTIONS

	MBE	Essay	PT
MBE	---	.733	.640
Essay	.939	---	.651
PT	.773	.891	---

Observed and corrected coefficients appear above and below the main diagonal, respectively.

The combination of MBE and essay scores accounted for .473 percent of the variance in PT scores.

The .94 corrected correlation between MBE and Essay scores suggests that if it was possible to make both tests (and particularly the Essay) much more reliable, then the relative standings of the applicants on the two measures would be almost identical. Moreover, if the MBE and Essay also were equally difficult, then they would make almost exactly the same pass/fail decisions. One implication of these findings is that observed differences in how the MBE and essay rank order applicants are probably far more a function of unreliability in the measures than any real differences in the underlying abilities assessed by these tests.

Effect of PT on Subgroups

Although the data in the total sample of 865 applicants indicates that the PT had virtually no influence on the percent passing the exam and only a small influence on who passed, it still may have had a large effect in certain subgroups. This issue was explored by examining the extent to which the PT would have made the same pass/fail decisions as the combination of MBE and Essay scores after controlling for differences in the relative difficulties of these sections.

The results of this analysis are presented in Table 8. These data indicate the following:

- o In all the groups, the PT made the same pass/fail decision as the combination of the MBE and Essay for about 75 percent of the applicants. The highest degree of agreement occurred with applicants from unaccredited schools and the lowest with women.
- o Women benefited slightly more than men from the inclusion of the PT (18.4% of them passed the PT but failed the combination of MBE and Essay whereas the reverse trend was true for 11.4%). The opposite pattern occurred for men: 9.8% passed the PT only and 13.5% failed the PT only.
- o There was a tendency for the PT to benefit Asian applicants more than Hispanic or Black applicants, however, there were too few applicants in any one group on which to base any firm conclusions regarding this matter. With only about 50 applicants per group, 5 applicants account for a 10% shift in passing rate! The same caveat holds for the results with graduates from unaccredited schools and attorney applicants.

The tendency for women to do better than men on the PT was most likely a by-product of a combination of two extraneous factors: (1) the total PT score is primarily dependent on the scores assigned to its written sections and (2) women tend to score higher than men on the written portions of the bar exam. This sex difference is shown in Table 9 where it can be seen that when the MBE and Essay were made equally difficult, 17.2 percent of the women passed the essay but failed the MBE whereas the reverse was true for only 6.2 percent. The opposite pattern occurred for men where 6.7 percent passed the essay but failed the MBE and 16.0 percent passed the MBE but failed the essay.

Differences in the relative effects of the PT among racial ethnic groups also may be explained by this "testing method" effect (i.e., some groups may do relatively better on essay type tests whereas other groups may do relatively better on multiple choice type tests). The presence of an interaction between group and testing method argues for the inclusion of both multiple choice and essay questions on a bar exam in order to help insure its fairness to all groups.

Attitudes Toward the PT

A questionnaire was mailed to all of the 865 applicants in the analysis sample shortly after the exam was administered. Applicants who did not respond to the first mailing were sent a reminder and another copy of the questionnaire. The results discussed below are based on the 664 applicants (77 percent of the analysis sample) who returned completed questionnaires prior to the announcement of who passed the exam.

There were no apparent major differences in the characteristics of the applicants who did and did not return questionnaires. For example, 51 percent of the applicants in both groups passed (after reappraisal) and their respective average exam scores were: MBE - 433 and 427; Essay - 403 and 399; and PT - 415 and 411.

Table 8

PERCENTAGE OF APPLICANTS WHOSE PASS/FAIL STATUS WAS AND WAS NOT CLASSIFIED IN THE SAME WAY BY THE PT AND THE COMBINATION OF MBE AND ESSAY SCORES WHEN THE SECTIONS WERE MADE EQUALLY DIFFICULT TO PASS

Group	Number of Applicants	Pass both	Fail both	Total agreed	Pass PT only	Fail PT only	Total disagreed
Total Sample	865	34.3	40.2	74.5	12.7	12.7	25.4
Males	570	32.5	44.2	76.7	9.8	13.5	23.3
Females	290	38.3	32.1	70.4	18.3	11.4	29.7
Asian	44	22.7	59.1	81.8	11.4	6.8	18.2
Hispanic	58	12.1	69.0	81.1	6.9	12.1	19.0
Black	54	5.6	66.7	72.3	9.3	18.5	27.8
Anglo	695	39.7	33.7	73.4	13.8	12.8	26.6
First Timers	656	42.8	30.0	72.8	13.6	13.6	27.2
Repeaters	209	7.7	72.3	80.0	10.0	10.0	20.0
ABA	549	45.7	27.5	73.2	14.4	12.4	16.8
Cal Accredited	180	11.1	63.9	75.0	10.6	14.4	25.0
Unaccredited	42	7.1	81.0	88.1	0.0	11.9	11.9
Attorney	34	17.7	61.8	79.5	2.9	17.7	20.6

Row percentages may not sum to 100.0% due to rounding and the number of applicants in each cluster may not sum to 865 because of missing data and some subgroups not having enough applicants to provide reliable data regarding overall agreement rates.

Table 9

PERCENTAGE OF APPLICANTS WHOSE PASS/FAIL STATUS WAS AND WAS NOT CLASSIFIED IN THE SAME WAY BY THE MBE AND ESSAY WHEN THESE SECTIONS WERE MADE EQUALLY DIFFICULT TO PASS

Group	Pass both	Fail both	Total agreed	Pass Essay only	Pass MBE only	Total disagreed
Males	36.7	40.7	77.4	6.7	16.0	22.7
Females	37.6	39.0	76.6	17.2	6.2	23.4
Total Sample	36.9	40.1	77.0	10.2	12.8	23.0

Row percentages may not sum to 100% due to rounding.

A copy of the questionnaire appears in Appendix A. A review of the questionnaire responses indicated the following:

- o Applicants felt actual legal practice helped them the most in preparing for the PT. Average scores on the five point scale from 1 = almost no help to 5 = a great deal of help were:
 - General law school courses = 2.4
 - Clinical courses = 3.0
 - Bar review courses = 3.1
 - Actual legal practice = 3.8
- o Applicants were slightly more apt to say the multiple choice portion of the PT helped than hurt them in answering the PT's written portion.
- o There was no consensus as to which part of the exam was the most difficult. Average scores on the scale from 1 = very easy to 5 = very hard were 3.5, 4.1, 3.6, and 4.5. for the Essay, MBE, and the AM and PM portions of the PT, respectively.
- o Most applicants felt the time limits were about right for the Essay and MBE, but too short for the PT. Average scores on the scale which asked about whether enough time was allowed (from 1 = far less than enough to 5 = far more than enough) were 2.8, 2.6, 2.1, and 1.6. for the Essay, MBE, and the AM and PM PT.
- o The especially low rating for the PM PT is consistent with it being rated the hardest part of the exam. Concern about the PT's time limits also was reflected by 42 percent of the respondents mentioning this problem in their written comments.
- o An equally high percentage wrote in favorable comments about the PT. Perhaps as a result of these different views, the overall reaction to continuing to include the PT as a regular part of the exam was essentially neutral (average rating of 3.2 on the scale from 1 = strongly opposed to 5 = strongly favor).
- o Applicants who earned high PT scores favored continuing the PT more often than applicants who earned low PT scores even though none of the respondents knew their scores or pass/fail status at the time they completed the questionnaire. This relationship was stronger than the one between an applicant's attitude towards continuing the PT and that applicant's MBE or Essay scores.

Applicants were asked to evaluate the realism of the case situations on each part of the exam and how well each part measured an applicant's legal knowledge and ability to perform as an attorney. Table 10 presents the average score on each of these dimensions. These data indicate the AM but not the PM PT received relatively high ratings for the realism of the case situations and how well they measured the ability to perform as an attorney. The low ratings of the PM PT probably stemmed from the previously noted feelings that not enough time was allowed to complete it. However, there was no difference between the AM and PM problems in average written scores (67.6 and 67.7) or in average multiple choice raw scores (11.271 and 11.275).

Table 10

AVERAGE RATINGS OF TEST CHARACTERISTICS

	Realism of case situations	Ability to perform as an attorney	Legal knowledge
Essay	3.1	2.8	3.4
MBE	2.5	2.1	3.1
AM PT	3.5	2.9	2.7
PM PT	3.0	2.5	2.2

All ratings were on a five point scale from 1 = very poor to 5 = very good

The questionnaire also asked applicants to evaluate the adequacy of various aspects of the PT's materials on a five point scale from 1 = not adequate to 5 = adequate. The respective average ratings for the AM and PM sections were: 3.4 and 2.8 for the file materials, 3.3 and 2.7 for the library materials, and 3.6 and 3.2 for the instructions. The higher ratings for the AM section are consistent with the ratings of the adequacy of the time limits on these sections.

Effect of Clinical Experience

The questionnaire asked applicants several questions about their prior legal experience, such as the number of credits they earned in clinical courses and whether they had ever been licensed to practice law. Table 11 shows that there was only a small relationship between such experience and PT scores that could not be explained by the skills already measured by the MBE and Essay. For instance, having ever held a job in a law office or clerked for a judge during the summer (item 4b) was positively correlated with PT scores, but it was also correlated to the same degree with MBE and Essay scores.

The one exception to this pattern was that applicants who felt that actual legal practice helped them prepare for the PT did better on it than applicants who either did not have such experience or felt it was not useful (item 8c). This finding suggests that some types of experience may indeed be helpful and others not; and, the questions we asked about practice experiences did not distinguish between these types.

The fact that attorneys who had been licensed in other states (or countries) did not do especially better on the PT than other applicants is an issue that deserves further study. It was noted that in general, attorney applicants do not do as well on either the Essay or the MBE as

Table 11

CORRELATIONS BETWEEN BAR EXAM SCORES AND QUESTIONNAIRE VARIABLES

Item number	Questionnaire Variable	MBE	Essay	PT
2	Number of credits earned in PT related courses in law school	.03	.03	.09
4a	Worked in a law office or clerked for a judge (versus marked no)	.17	.19	.25
4b	Worked in a law office or clerked for a judge during the summer	.30	.32	.34
4c	Worked in a law office or clerked for a judge during the school year	.10	.10	.15
6b	Had been licensed to practice law in the U.S prior to July 1983	.11	.02	.09
8c	Felt that actual legal practice helped in preparing to take the PT	.02	-.01	.16
13c	Adequate time to answer AM PT	.03	.00	.10
13d	Adequate time to answer PM PT	.09	.01	.08
18	Would favor continuing the PT on future bar exams	.10	.05	.20
19	Wrote a favorable comment about the PT	.04	.10	.18

first time takers from ABA schools (see Table 8). However, the passing rate (47%) among all of the 126 attorney applicants who took the July 1983 exam was 21% higher than it was for this group on the July 1982 exam. It is not known whether this increase was due in part to the PT.

Correlates of Bar Exam Scores

Table 12 contains the correlations of bar exam scores with several background characteristics, including both school reported and self reported Law School Admissions Test (LSAT) scores. These data indicate that self reported class standing correlated slightly higher with Essay scores than with PT scores. There was a .63 correlation between school reported LSAT scores and total bar exam scores.

There was a very high correlation (.90) between self reported and school reported LSAT scores. In other words, the self reported LSAT scores rank ordered applicants in much the same way as the LSAT scores reported for them by their law schools. However, a self reported score was, on the average, 18.5 points higher than a school reported score. Self reported class standing had a relatively low correlation with both self and school reported LSAT scores (.23 and .17, respectively).

SUMMARY AND CONCLUSIONS

The July 1983 version of California's bar exam contained a six question Essay test, the Multistate Bar Examination (MBE), and for the first time, two Performance Tests (PTs). The morning PT involved a torts case and the afternoon PT a civil procedures case. Unlike the typical essay question, each PT contained a file of statutes, case law, memos, and other materials regarding a hypothetical client's case. Applicants were asked 15 multiple choice questions about these materials. They also were instructed to use them in preparing a memorandum that would be used in handling an important step in the case.

An analysis of the effect of adding the PT to the bar exam was conducted with a random sample of 865 of the 7,277 applicants who took all three parts of the exam. For the purposes of this research, all 865 of these applicants had all of their essay and PT written answers read twice during the regular grading process. About 77 percent of the applicants in this sample completed a questionnaire (prior to the release of exam scores) that inquired about their clinical legal experience and their attitudes towards the various sections of the exam.

The major findings of this research were as follows:

- o PT readers agreed with each other in the score that should be assigned to an answer to about the same degree that essay readers agreed with each other in the scores they assigned.

Table 12

CORRELATIONS BETWEEN BAR EXAM SCORES
AND BACKGROUND CHARACTERISTICS

	MBE	Essay	PT	Total score
Class standing*	.38	.47	.34	.44
LSAT score*	.47	.38	.46	.51
School reported LSAT score	.60	.47	.55	.63
Racial/ethnic group = Anglo*	.27	.27	.28	.31
Sex = Female*	-.06	.12	.15	.07
ABA graduate	.36	.33	.41	.42
First Timer	.33	.32	.39	.39

* Self reported

- o The total PT score (morning plus afternoon sessions) was almost as reliable (.69) as the Essay score (.77). The PT's written score was almost as reliable (.67) as its total score (.77). Thus, the multiple choice portion of the PT did not substantially increase the reliability of the total PT score. Moreover, the written score correlated .96 with the PT total score. These findings suggest that a substantial amount of testing time could be saved without losing a significant amount of reliability if the multiple choice portions of the PT were eliminated.
- o Applicants tended to earn much higher MBE than Essay scores. Because the difficulty of the PT fell in between the MBE and the Essay and because all three sections of the exam carried about equal weight in determining an applicant's total score, the inclusion of the PT did not influence the percent passing.
- o The inclusion of the PT affected the pass/fail status of about 8 percent of the applicants. However, some of this effect may be due to the less than perfect reliability of the scores on all three sections of the exam.
- o The passing rate in different racial/ethnic groups was not appreciably affected by the inclusion of the PT on the exam.
- o Women applicants tended to do slightly better on the PT than would have been expected given their total scores on the other two sections, however, this trend appeared to be due to the fact that women tend to score relatively higher on essay than on multiple choice questions and two-thirds of the PT total score involved a written response.
- o Self reported and school reported LSAT scores as well as self reported class standings correlated about as well with PT scores as they did with MBE and Essay scores. The correlations of school reported LSAT scores with bar exam scores was quite strong considering the LSAT scores were earned almost four years before the applicants took the bar exam.
- o The major complaint about the PT was that not enough time was allowed to complete it. This was particularly true for the afternoon session. However, afternoon PT scores were just as highly correlated with other variables as were the morning PT scores.
- o The realism of the PT's case situations was judged to be much higher than that of the MBE's or Essay's situations. The PT was rated about the same as the Essay but better than the MBE in terms of how well it tested an applicant's ability to perform as an attorney.
- o There was substantial variation among applicants in their opinions about whether the PT should be included on future examinations. Applicants who favored continuing it on the exam tended to earn higher PT scores than those who were opposed to its continued use.

- o Applicants felt that actual legal practice helped them the most in preparing for the PT. And, applicants who said that such practice helped, actually did better on the PT than would have been expected given their scores on the other sections of the exam. There did not appear to be any other significant relationship between PT scores and prior experience that could not be explained by the combination of the applicants MBE and Essay scores.

Several factors may have influenced the findings reported above. Some of these factors are: (1) all the comparisons among sections in the results portion of this report were conducted after making the three sections equally difficult to pass rather than reflecting the observed differences in their difficulties; (2) applicants (and bar review courses) were informed the July 1983 exam would include the PT (and provided with a prototype copy of it) about seven months before it was administered (thereby reducing the possible effect of special preparation for it); (3) the two PTs were very similar in some respects but very different in others; (4) future PTs may differ in a number of ways from those given in July 1983; and (5) the size of the effect of including the PT on who passes the exam is affected by the amount of weight assigned to it in computing an applicant's total score, the specific pass/fail rules employed, the scoring standards employed on it, and the overall percent passing. For example, if the passing rate on the California exam was 80 percent or higher and if all three sections carried equal weight and were equally difficult, it is highly unlikely that as many as 8 percent of the applicants would have had their status changed by the inclusion of the PT.

In light of these considerations, the reader is cautioned that the effect of the PT on the July 1983 results may not be indicative of the PT's impact on subsequent California exams or on the exams given by other states should they decide to include a PT on their bar exam.

REFERENCES

- 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, 1981.
- Klein, S. An analysis of the July 1983 California Bar Examination. Report prepared for the Committee of Bar Examiners of the State Bar of California, 1984.

APPENDIX A

APPLICANT QUESTIONNAIRE AND PERCENTAGE SELECTING EACH CHOICE

Dear Candidate:

The Committee of Bar Examiners of the State Bar of California and the National Conference of Bar Examiners are conducting an in depth analysis of the performance test that was administered as part of the July 1983 California State Bar Examination. This analysis is being conducted on a small percentage of the applicants who took this exam. Your name was selected randomly to participate in this study.

The purposes of this study are to identify ways of improving the bar examination and to explore what background factors, if any, are related to performance on it. Your responses will therefore have a significant impact on decisions about the nature of future examinations.

Your answers to the following questionnaire will be treated in the strictest confidence and will not in any way affect the grading of your answers. The data will be used solely for research purposes.

In return for your cooperation in this research, you will receive a copy of the summary report of our findings.

Thank you for taking the time to consider this request and I hope you will be willing to participate in this important research. If you have any questions about the questionnaire or the study, please write to me at the address above.

If you did not take the Performance Tests, please check this box and return this form. If you did take it, please answer the following questions and return your questionnaire by September 30. A return envelope is enclosed for your convenience. Thank you.

Sincerely,

Stephen P. Klein, Ph.D.

1) To the best of your recollection, what was your LSAT score? If you took the LSAT more than once, please indicate your highest score. If you did not take the LSAT or have no recollection of your approximate score, mark an X ... (Mean = 601, SD = 83.5, N = 522)

2) While you were in law school, how many credits did you earn in courses that required performing practice oriented tasks, such as conducting legal research, preparing motions or briefs, examining witnesses, and interviewing clients? If you did not take any courses that involved these activities, skip to Question 4 (Mean = 7.8, SD = 4.4, N = 612)

3) About what percentage of the time spent in these courses was devoted to SIMULATED cases and what percentage to REAL cases?

a) SIMULATED CASES, such as moot court or an on-campus clinical program..... (Mean = 25.7% among 564 respondents)

b) REAL CASES where you had contact with the client or the client's attorney, such as part of a on-campus or off-campus clinic run by your law school.... (Mean = 41.7% among 307 respondents)

4) Have you ever worked in a law office (or clerked for a judge) where you had to use your legal knowledge and skills in handling actual cases, such as conducting legal research, preparing motions or briefs, and interviewing clients? Circle each period of employment that lasted at least 4 weeks and was not part of a law school program.

- 20% a) No
- 48% b) Yes, during the summer
- 57% c) Yes, during the school year: average number of hours/week (Mdn = 20) ...
- 33% d) Yes, after graduation: average number of hours/week (Mdn = 40)

If you answered "yes" to question 4, please circle whether you:

- 66% a) were paid for your services
- 8% b) did the work as a volunteer (N = 523 said Yes)
- 26% c) both a and b

5) About where did you rank in your graduating law school class?

- 7% a) bottom 20%
- 0% b) next 20%
- 33% c) middle 20%
- 32% d) next 20%
- 28% e) top 20%

6) Have you ever been licensed to practice law?

- 90% a) No
- 8% b) Yes; number of years practiced in the U.S. prior to July 1983
- 3% b) Yes; number of years practiced outside the U.S. prior to July 1983

Range of years licensed to practiced = 1 to 33, Mdn = 3

7) Which one of the following schedules would you recommend for future examinations?

- 54% a) Give the 6 essay questions on day 1 and the 2 performance test problems on day 3...plan used for July 1983 examination.
- 4% b) Give the 6 essay questions on day 3 and the 2 performance test problems on day 1
- 25% c) Give 3 essay questions and 1 performance test problem on day 1; give the remaining 3 essay questions and performance test problem on day 3.
- 16% d) Lengthen the time allowed to answer each performance test problem by adding a fourth day to the examination.

8) Circle one choice for each of the activities below to indicate how much it helped you to prepare for the Performance Test:

	Almost no help at all 1	2	Some help 3	4	A great deal of help 5	
a. general law school courses	30	19	33	12		6%
b. clinical courses	17	16	31	23		13%
c. actual legal practice	8	8	21	22		21%
d. bar review course(s)	18	15	29	19		19%
e. other _____						

9) In your opinion, how difficult was each part of the exam? Please circle one number for each part. The AM performance test, Carelton v. Mid-Central Shipping, was a Torts case involving a boy who was injured while jumping off a plank. The PM performance test, Lee v. Western Steel, was a Civil Procedure case involving a class action suit dealing with acid fog.

	Very easy 1	Easy 2	Average 3	Hard 4	Very hard 5	
a. Essay	1	4	49	38		8%
b. Multistate (MBE)	0	2	23	39		36%
c. AM Performance Test - Carelton	1	6	35	41		17%
d. PM Performance Test - Lee	0	2	10	26		62%

10) In your opinion, how good a measure of your LEGAL KNOWLEDGE was each part of the examination? Please circle one number for each part.

	Very poor 1	Poor 2	Fair 3	Good 4	Very good 5
a. Essay	5	12	31	40	11%
b. Multistate (MBE)	11	19	31	29	10%
c. AM Performance Test - Carelton	11	28	30	21	4%
d. PM Performance Test - Lee	35	31	20	10	11%

11) In your opinion, how good a measure of your ABILITY TO PERFORM AS AN ATTORNEY was each part of the exam? Circle one number for each part.

	Very poor 1	Poor 2	Fair 3	Good 4	Very good 5
a. Essay	16	20	37	23	4%
b. Multistate (MBE)	32	32	26	7	2%
c. AM Performance Test - Carelton	18	19	25	29	9%
d. PM Performance Test - Lee	30	23	19	20	7%

12) Circle one number for each part of the exam to indicate your opinion of the REALISM OF ITS CASE SITUATIONS:

	Very poor 1	Poor 2	Fair 3	Good 4	Very good 5
a. Essay	7	16	44	29	5%
b. Multistate (MBE)	20	28	36	14	3%
c. AM Performance Test - Carelton	5	7	34	40	15%
d. PM Performance Test - Lee	15	15	32	29	9%

13) Circle one number for each part of the exam to indicate your opinion of whether ENOUGH TIME WAS ALLOWED to answer it:

	Far less than enough 1	2	About right 3	4	Far more than enough 5
a. Essay	3	18	72	7	1%
b. Multistate (MBE)	10	27	56	5	1%
c. AM Performance Test - Carelton	33	31	33	3	1%
d. PM Performance Test - Lee	61	23	14	2	0%

14) Which one of the following statements best reflects your opinion of the multiple choice portions of the performance test?

- 9% a. Led me off in the wrong direction for the written portion
- 53% b. Did not help or hurt me in preparing for the written portion
- 38% c. Helped me prepare for the written portion

15) Circle one number for each performance test to indicate the ADEQUACY OF THE FILE MATERIALS for preparing your written answer (memo):

	Not adequate				Very adequate
	1	2	3	4	5
a. AM Performance Test - Carelton	4	12	39	34	11%
b. PM Performance Test - Lee	15	22	35	20	8%

16) Circle one number for each performance test to indicate the ADEQUACY OF THE LIBRARY MATERIALS for preparing your written answer (memo):

	Not adequate				Very adequate
	1	2	3	4	5
a. AM Performance Test - Carelton	5	15	35	32	13%
b. PM Performance Test - Lee	23	22	28	18	10%

17) Circle one number for each performance test to indicate the ADEQUACY OF ITS INSTRUCTIONS:

	Not adequate				Very adequate
	1	2	3	4	5
a. AM Performance Test - Carelton	4	7	31	36	21%
b. PM Performance Test - Lee	12	14	30	27	17%

18) What is your opinion regarding whether the performance test should continue to be a regular part of the bar exam?

- 16% a. strongly opposed
- 18% b. opposed
- 14% c. neutral
- 37% d. favor
- 15% e. strongly favor

19) What were your general reactions to the Performance Test? How could it be improved? What parts should be maintained, eliminated, and changed? Please put your comments on the reverse side of this form.