

RESEARCH ON THE CALIFORNIA BAR EXAMINATION:  
A TEN YEAR RETROSPECTIVE

A report prepared for  
The Committee of Bar Examiners  
of the State Bar of California

Prepared by  
Stephen P. Klein, Ph.D.  
GANSK & ASSOCIATES

December 22, 1982



RESEARCH ON THE CALIFORNIA BAR EXAMINATION:  
A TEN YEAR RETROSPECTIVE

This report summarizes the major findings of the studies conducted on California's bar examination during the past 10 years. These studies investigated: various characteristics of the examination (such as the appropriateness of its time limits), essay grading practices, the accuracy of multiphased grading, the stability of pass/fail standards, correlates of initial and eventual pass/fail status, possible racial/ethnic and sex biases, and the relationship of traditional measures of legal skills and knowledge to the ability to perform legal tasks. A list of the studies conducted is attached at the end of this report, and hereinafter the studies are referred to by number as shown on the list, e.g. (4, 6).

CHARACTERISTICS OF THE BAR EXAMINATION

California's General Bar Examination (GBX) contains a one day multiple choice section and a two day essay section. The multiple choice section is the Multistate Bar Examination (MBE). The MBE consists of 200 items that are drawn from 6 content areas. The essay section currently contains 9 questions. Applicants can pass the GBX by earning a combined total score (i.e., MBE + essay) equivalent to 70% of the maximum possible total score and/or by passing one section on one administration and the other section on another administration.

Research on the GBX indicates: there is a strong correlation between essay and MBE scores (e.g., applicants that pass one section are likely to pass the other section whereas those who fail one section are likely to fail the other section); about 16% more first timers pass the MBE as pass the essay; even substantial increases in the time allowed to answer an MBE or essay question does not change an applicant's relative standing any more than would be expected by chance; and the sequence in which the questions are asked does not affect an applicant's performance on these questions (4, 6, 11, 13).

ESSAY GRADING PRACTICES

Possible inconsistencies in grading an essay question were investigated by having several readers independently score a common set of answers as well having a reader grade the same set of answers on two or more occasions (2, 8, 11). Results of this research indicate: readers agree with themselves only slightly better than they agree with each other; using an analytic (scorecard) grading system is not as cost effective as the traditional (holistic) method; having applicants answer about 9 questions balances out most of the inconsistencies that can be avoided by the use of multiple questions; and using the average of two independent readings of an answer also helps to balance out reader inconsistencies (but not as much as having multiple questions).

The grading practices research and other studies (6, 9) show that the total score on a 9 question essay test, by itself, is not reliable enough to make pass/fail decisions about individual applicants if each answer is read only once. This would be so even if there was as many as 20 essay questions. However, the current combination of 9 essay questions plus the MBE provides a total score that is sufficiently reliable for making pass/fail decisions.



## MULTIPHASED GRADING

The bar examination is used to make pass/fail decisions about individual applicants rather than to determine by how much an applicant passed or failed. Research indicated these decisions could be made with greater accuracy if reader resources were concentrated on those applicants whose pass/fail status was most in doubt (2); and, such reallocation would not introduce other errors (4, 5). These findings led to a policy of passing applicants if they had sufficiently high scores on a combination of the MBE and 3 essay questions. Pass/fail decisions on the remaining applicants are made after reading all 9 of their essay answers at least once, and for those near the pass/fail line, two or more times.

Tests of the multiphased grading system show that: it reduces the total number of answers that have to be read; it increases the reliability of pass/fail decisions; it increases the passing rate by about 0.5% (i.e., compared to the rate if all applicants had all of their answers read at least once); and its accuracy is not affected by which 3 essay questions are read in the first phase of the grading process (7, 11).

## STABILITY OF PASS/FAIL STANDARDS

A different set of questions is asked each time the GBX is administered. Thus, the percent passing a given administration could be affected by any variation in the GBX's difficulty across administrations as well as by the ability of the group taking that examination.

On the multiple choice portion of the GBX, raw scores (i.e., the number of questions answered correctly) are converted to scale scores in order to adjust for possible variations in average question difficulty across administrations. Empirical studies with the MBE have demonstrated the very high degree of accuracy of the standard statistical procedures that are used for this scaling (18).

It is not possible to adjust the essay scores in the same way the MBE scores are adjusted. Nevertheless, studies of past California examinations indicate that essay scores have not been biased by possible variations in question difficulty (and/or essay grading standards) across administrations (4, 9, 11). Thus, differences in percent passing from one examination to another stem from differences in the applicants' legal skills and abilities rather than from variations in test difficulty.

## CORRELATES OF GBX SCORES

Applicants with English and science undergraduate majors have higher average GBX scores than applicants who majored in the social sciences or education (12). The only exception to this trend is that male, but not female, science majors score below average on the essay section.

GBX scores are highly correlated with law school admissions scores and law school grades (1, 3, 6). In fact, the correlation is about four times stronger than it is between law school grades and the combination of undergraduate grades and law school admissions scores. The essay portion of the GBX tends to be more highly correlated with law school grades than with admissions scores while the opposite is true for the MBE.



The first year law student's examination (FYLSX) is a good predictor of an applicant's initial score and eventual pass/fail status on the GBX (10). Applicants who score at the pass/fail line on the FYLSX are likely to fail on their first attempt to pass the GBX. However, they have a slightly better than 50/50 chance of eventually passing the GBX.

In general, applicants from American Bar Association (ABA) approved law schools earn higher GBX scores than applicants from schools that are only accredited by California who in turn receive higher scores than applicants from unaccredited schools (6, 11, 13). However, there are numerous exceptions to these trends. In fact, from time to time, an unaccredited school will have a higher pass rate than an ABA approved school. And, there is less difference among school types in eventual passing rates than there is among them on any given examination (10).

Applicants of a given ability level (as indicated by their law school admissions scores) are more likely to pass the GBX if they graduated from an ABA school than if they graduated from some other type of law school. A few ABA schools have especially high passing rates relative to the ability level of their graduates (3, 6).

First time takers have a much higher passing rate than repeaters (6, 13). Repeaters who came close to passing on their first attempt are far more likely to eventually pass (and in fewer attempts) than are repeaters who score well below passing on their first try (10). However, there is no initial score that clearly distinguishes between the repeaters who will versus will not eventually pass. About 5% of the repeaters (or 1% of all applicants) pass the GBX as a result of the rule that permits passing the GBX by passing one section on one administration and the other section on another administration. However, far more than 5% of the repeaters fail because they try to pass solely on the basis of this bifurcation rule.

#### RACIAL/ETHNIC BIAS

The percentage of minority applicants that pass the GBX is well below the percentage of Anglo applicants that pass. This gap has occurred despite the fact that about 43% of all the minority applicants taking the GBX graduated from the 5 ABA schools with the highest average admissions scores (6). Only 23% of the Anglo applicants graduated from these 5 schools.

Three types of studies were conducted to determine whether the differences in passing rates among four racial/ethnic groups (Anglo, Asian, Black, and Hispanic) were attributable to certain characteristics of the examination or to differences in the average academic ability level of the groups. The three types of studies were: grader bias, item bias, and test bias.

The grader bias study (1) showed that an applicant's score on an essay answer was not affected by whether or not the reader of that answer was a member of the same racial/ethnic group as the applicant.

The item bias studies (1, 17) found that various types of multiple choice or essay questions were not especially difficult or easy for minority groups. In other words, the relative differences between groups were not affected by the type of question asked or the content area covered by a question.





The test bias studies (1, 3, 6) indicate that the differences in average score between groups on both the essay and multiple choice portions of the GBX are consistent with the differences in their law school admissions scores and law school grades. Thus, the GBX does not widen or narrow the gap in performance level between groups. The gap also would not change if the GBX's time limits were substantially increased (11, 13) or if the GBX were expanded to include oral and written clinical skills tests (14, 15).

#### SEX BIAS

Women generally have higher GBX scores than men. They also tend to have higher law school admissions scores and grades. The sex bias studies (3, 6) show that after controlling for differences in law school admissions scores and grades, women tend to do slightly better than men on the essay portion of the GBX whereas men tend to do slightly better than women on the multiple choice portion. However, these differences are not large enough to appreciably affect the relative passing rates of the groups.

#### RELATIONSHIPS WITH PERFORMANCE TESTS OF LAWYERING SKILLS

Three performance tests were given in conjunction with the July 1980 administration of the GBX. One test assessed some of the skills that are required for carrying out legal research, such as the ability to determine whether and how decisions in the legal literature could be used to support a client's case (14). Applicants taking the research test received copies of the statutes, cases, and other relevant documents on which they were to base their answers. Thus, it was similar to an open book examination.

The second performance test assessed certain trial practice skills (16). In this test, applicants were given background materials about a case, shown a brief segment of the legal proceedings via a videotape, and then asked 1 to 3 questions about the segment, such as whether an objection that was made should be granted. Applicants had 5 to 10 minutes to answer the questions before the next segment in the proceedings appeared on the screen.

The third test was administered via an Assessment Center that was conducted two weeks after the GBX (15). An applicant participated in the center for two days. On one day, the applicant functioned as the attorney for the plaintiff from the beginning to the end of a simulated case. On the other day, the applicant served as defense counsel in a different case. Each day, the applicant took several oral and written tasks. Specially trained actors played the roles of clients and witnesses for the oral tasks. Performance on an oral task was videotaped for later evaluation.

Analyses of the performance test data indicate they can be administered under standardized conditions and scored reliably. Applicants report these tests are better and more realistic measures of their legal skills than are either the essay or MBE portions of the GBX. Applicants with some clinical legal experience score higher on the performance tests than do applicants without such experience. Research, Trial Practices, and Assessment Center scores correlate about as well with GBX scores as MBE and essay scores correlate with each other. Scores on the performance tests correlate higher with each other than they do with GBX scores.



In summary, the performance tests measured a related group of skills; and, the skills in this group are similar to, but not exactly the same as, those measured by the GBX. Cost, test security, and other factors suggest that some of the machine scorable and essay portions of the performance tests, but not the oral portions, could be included in a bar examination.

#### REFERENCES

Reports prepared for the Committee of Bar Examiners:

- 1) An investigation of possible item and grader biases in a state bar examination. (1976)
- 2) An analysis of grading practices on the California Bar Examination. (1977)
- 3) An analysis of possible sex and racial/ethnic biases in the July 1976 California State Bar Examination. (1978)
- 4) Summary of studies conducted for the Committee of Bar Examiners and the statistical rationale underlying proposed changes. (1978)
- 5) Comparison of an 8, 9, and 12 answer essay test in a multiphased pass/fail decision model. (1978)
- 6) An analysis of the relationships between bar examination scores and an applicant's law school, admissions test scores, grades, sex, and racial/ethnic group. (1979)
- 7) A comparison of the effectiveness of a single versus multiphased grading system. (1980)
- 8) Intra- and inter-reader agreement on the essay section of the California State Bar Examination. (1980)
- 9) An analysis of possible variations in pass/fail standards on the California State Bar Examination. (1981)
- 10) The relationship between initial score and eventual pass/fail status on the California Bar Examination. (1982)
- 11) Analysis of the February 1982 bar examination. (1982)
- 12) The relationship between undergraduate major and bar examination scores. (1982)



Reports prepared for the Committee of Bar Examiners and the National Conference of Bar Examiners:

- 13) The effect of time limits, item sequence, and question format on applicant performance on the California Bar Examination. (1981)
- 14) Testing research skills on the California Bar Examination. (1981)
- 15) The relationship between clinical skills and bar examination results. (1982)
- 16) The relationship between trial practice skills and bar examination results. (1982)

Reports prepared for the National Conference of Bar Examiners involving California data:

- 17) An evaluation of the Multistate Bar Examination. (1982)
- 18) Technical Manual for the Multistate Bar Examination. (in preparation)

