Title

An Examination of Possible Item, Test, and Grader Bias in the California Bar Examination

Permalink

https://escholarship.org/uc/item/9k94r7nk

Journal

National Black Law Journal, 4(3)

Authors

Klein, Stephen P.
McDermott, Anthony

Publication Date

1975

Copyright Information

Peer reviewed

AN EXAMINATION OF POSSIBLE ITEM, TEST, AND GRADER BIAS IN THE CALIFORNIA BAR EXAMINATION

Stephen P. Klein
University of California, Los Angeles
and
Anthony McDermott
University of Oregon

This report describes a series of studies to determine if there was cultural bias in the California Bar Examination. Two of the studies, the Item Bias Study and the Test Bias Study, were done in conjunction with the July, 1973 Bar Examination. The Test Bias Study included within it a component to determine the effect upon Bar performance of Background and Attitude factors. The third study, the Grading Practices Study, was done in conjunction with the July, 1974 Bar Examination.

In order to identify Black, Chicano and Anglo candidates who took the July 1973 Examination and to gather data concerning socioeconomic status and attitudes toward the Bar Examination, EEA sent out a questionnaire to all persons who took that Examination. From the pool of returned questionnaires a group of 128 Blacks and 83 Chicanos who had graduated from California Law Schools was identified. A sample of 342 Anglo graduates of California Law Schools was randomly selected from the same pool of returned questionnaires. The law schools for each of the 553 candidates were then contacted in order to obtain needed data concerning the candidates. As a result of numerous factors ranging from a candidate not taking an admission's test to missing school records, complete data was available for 248 Anglos, 83 Blacks and 52 Chicanos.

ITEM BIAS STUDY

The Item Bias Study was designed to examine whether certain questions (items) or kinds of questions were unduly difficult or otherwise inappropriate for minority group members. In order to identify such items the average performance of the Black sample and the Chicano sample was measured against the average performance of the Anglo sample. If on certain items the difference between the average performance of the Blacks and/or Chicanos and the Anglos was significantly greater (or less) than the difference between the average performance of the groups on the whole Examination then these items could be deemed biased against (or in favor of) the minority groups. (It should be noted that the Anglo sample had a passage rate on the Examination of over 50% while the Black and Chicano samples had a passage rate of less than 25%.) The Item Bias Study tested the fairness

as defined above of each of the 20 essay questions and each of the 200 MBE questions as well as the fairness of each of the five MBE subtests. Thus the Item Bias Study had three component parts to it. The results as to each of the components are as follows:

(1) *MBE Subtests* - The determination of whether the subtests per se are fair (i.e. as opposed to the MBE test as a whole or particular items in it) involved an examination of the pattern of average performance on each subtest for each ethnic group. For example, was the size of the difference between average Anglo and Chicano scores on Torts significantly larger or smaller than the size of the difference on Contracts?

The results of this analysis indicated that there was statistically significant difference in the pattern of scores across the five subtests. However, although a statistically significant difference was found, the actual practical magnitude of the difference was quite small; somewhat less than 1% of a candidate's score on a given subject could be attributed to the pattern differences between ethnic groups on that subtest.

(2) *MBE Items* - The issue of item fairness involved determining whether or not there were any consistent differences between groups in their pattern of performance across items. The results of this analysis indicated that there is no practical evidence of any bias in the MBE due to certain items being atypically more or less difficult for minority groups. In other words the magnitude of the difference between Anglo and minority group performance is relatively constant across all items and what deviations are observed are essentially due to chance fluctuations rather than any systematic interaction between items and ethnic groups.

(3) *Essay Items* - The procedure used to examine the fairness of the essay items was comparable to that used with the MBE items except for the following changes: (1) since all candidates did not answer all essay questions and since the analysis procedure required no missing data, where a candidate did not answer a question he/she was assigned to the average score for his/her ethnic group on that question; and, (2) an additional analysis was run to determine whether there were any systematic differences between groups in which questions they chose to answer.

The results of this part of the study indicated statistically but not practically significant differences between groups in their respective patterns of performance across the questions. Also the results of the analysis in which questions each group chose to answer did reveal certain differences between groups. In other words, the proportion of students who chose to answer a given question in one group was occasionally quite different than the proportion choosing to answer this question in another group. However, although there was no correspondence between these choice differences and the relative difficulties of the question, one pattern was noted: In each of the four sessions a much greater percentage of Anglos chose to answer the optional question than did Blacks or Chicanos.

TEST BIAS STUDY

This study was based on the premise that unfairness in an examination may be pervasive rather than centered in certain items or sets of items. The

usual methodology for inquiring into test bias involves two steps. First compute an equation that describes the typical relationship for all candidates between a predictor (or a set of predictors) and some independent criterion of success. A typical example of this is the relationship between college entrance examination scores (a predictor) and grades in college (a criterion). Then determine whether a given group's actual criterion scores are generally higher or lower than what would be predicted by the equation developed in the first step. If the group's subsequent performance on the criterion was "under" predicted to a significant degree the predictor would be biased against the group in the sense that the predictor underestimated the group's true potential to achieve on the criterion.

Since there is no available criterion of subsequent performance, i.e., there is no commonly accepted measure of legal ability after passing the Examination and being admitted to practice, the approach described above for measuring bias on the Bar Examination could not be utilized directly. However the logic of that approach was utilized and adapted to fit the uniqueness of the Bar Examination's situation. This was done by examining whether minority group performance on the Bar Examination is generally over or under predicted by presumably unbiased indicators of how well a person will perform on the Bar. The rationale underlying this new approach is as follows: If an Anglo and a Black have the same predictor score, such as Law School Grade Point Average (LGPA), they should do equally well on the Bar. If the Anglo does better on the Bar Examination, there is reason to believe that the Examination is not fair to the minority student.

An equation was thus computed for each ethnic group relating three predictors of bar performance, Undergraduate Grade Point Average (UGPA), Law School Admission Test score (LSAT) and Law School Grade Point Average (LGPA) with three performance criteria, the MBE and Essay subtest scores and the total Bar Examination score. These equations in each instance indicated that if an Anglo and a Black or a Chicano candidate had the same predictor score, the Anglo would generally do better on the Bar. This would suggest that the Bar is biased against minority candidates. A further analysis indicated that although the magnitude of the bias is not the same throughout the distribution of predictor scores, for example the bias appears greatest for Blacks with low LSAT scores while the bias against Chicanos is greatest for those with high LSAT scores, still at no time is the Examination biased in favor of either group.

These findings mean that if an Anglo and a minority candidate had the same predictor score, the Anglo would generally do better on the Bar. In short, it would suggest that the Bar is biased against minority candidates. *However the groups do not have comparable predictor scores. Further the real issue is not what score you receive on the Examination, but rather, whether or not you pass it.* Thus the seriousness of the bias in the Examination is a function of how many minority candidates actually passed the Examination relative to the number that would have passed had no bias existed.

A determination of how many Blacks and Chicanos would have passed the Bar had no bias existed was computed in two different ways. Under the first method minority candidates were given bonus points comparable to the

difference between their bar scores and the comparable bar score of an Anglo with the same predictor. For example, a Black candidate with an LSAT score of 541 would be expected to score 1547 on the Bar, whereas an Anglo with a 541 LSAT would be expected to score 1600. Thus a Black with a 541 LSAT would receive 53 bonus points. *The results of this method indicated no significant difference between the percentage of minority group members who actually passed versus those who would have been expected to pass had no bias existed.* The most apparent reason for this finding is that the minority group bar scores were so low, that even with the bonus, few candidates had an expected bar score greater than the 1600 points necessary to pass the Exam. In fact this method of analysis indicated that more minority group members passed than would be expected relative to their adjusted predictor scores and to that extent it indicated a slight bias in favor of minority group members.

The second method for determining how well minority group members would have done on the Bar had no bias existed was to determine what score on a predictor projected a passing Bar score for Anglos. In the case of the LSAT for instance this was found to be a 541 in that an Anglo with a 541 LSAT score could be expected to attain the 1600 points necessary to pass the Bar. Then given the average LSAT score for the minority group candidates and using a normal probability curve, the number of minority group applicants who would likely obtain a score of 541 or better was computed. The number of minority candidates who could be expected to score over 541 on the LSAT then represented the number of minority candidates who could be expected to pass the test. The number of candidates in each group who could be expected to pass the test was then compared to the actual number passing in each group. *The results of analyses using all three predictors were identical to those in Method #1 in that there was no evidence that more minority applicants would be expected to pass the Bar than actually did pass the Bar.*

Thus, both methods #1 and #2 indicated that whatever bias against minority groups existed on the Bar Examination (or its two major subsections), it was not serious enough to influence the proportion of minority group members who passed versus those who did not pass the examination. In other words, the marked disproportionability in the number of Anglo versus minority applicants passing the bar is explained more by differences in ability than by any systematic bias against minority members within the examination itself.

The assumption underlying the analysis of test bias is that the predictors of bar performance (e.g., LSAT and LGPA) are themselves unbiased with respect to assessing a candidate's ability. Although bias might exist in LGPA, it is not likely in most law schools since special precautions are often taken to guard against this problem (such as examination graders not knowing the identify of the students whose papers they grade). If bias does exist in LGPA's, then one must assess its direction; i.e., does it tend to favor or penalize the minority student? For example, if a minority student tended to take harder courses than Anglo students, then the minority student's LGPA would underestimate that student's true performance level. If, on the other hand, the minority student took courses that tended to be graded leniently, then their LGPA's would be biased in their favor. At the present time, there

is no evidence regarding the relative proportions of Anglo and minority students taking courses that differ in difficulty. The Grading Practices Study deals with possible bias in the scoring of examination papers much like those in law school and, thus, the results of that study will be relevant to the issue of possible grader biases influencing LGPA's.

Possible bias in LSAT scores was examined by determining the typical relationship between LSAT and LGPA for each ethnic group. The rationale underlying this analysis is that persons with equal LSAT scores should do equally well in law school. If the minority group member does better than the LSAT is said to be biased against that person in the sense that the LSAT underestimated his her true potential. If on the other hand, the minority group member does less well than the Anglo with the same LSAT score, then the LSAT would be biased in favor of the minority group member, i.e., it overestimates that applicant's true potential. To carry out this analysis, regression equations were computed for predicting LGPA's from LSAT scores for each ethnic group. On the basis of these equations, it was determined that for an LSAT score of 500, the typical LGPA's were 2.61, 2.32 and 2.29 for Anglos, Blacks and Chicanos, respectively. It appears, therefore, that the LSAT may be biased in favor of minority groups in the sense that compared to Anglos, it overestimates minority group performance levels. This finding may account for at least some of the discrepancy in Bar performance of Anglo and minority candidates with comparable LSAT scores. In other words, since the LSAT appears biased in favor of the minority candidates, then one would predict that for comparable LSAT scores, the Anglo candidate would tend to perform better on the Bar than his/her minority colleague.

It should be noted that the present study was not designed to determine whether the LSAT was biased and examined this issue only to assess the plausibility of the assumption that it was "fair" to the minority candidates involved in this research. Further, the present study was limited to just law school graduates from one state. The findings in this research are consistent, however, with those of Schrader and Pitcher (1973, 1974) who found that the LSAT overpredicted the first year law grades of Blacks and Chicanos; i.e., the LSAT was biased in their favor in every law school sampled.

On the basis of the foregoing discussion and findings, it seems reasonable to conclude that there is no statistical evidence that the predictors of Bar performance are biased against minority group members. In fact, with respect to LSAT, there is a strong evidence indicating that just the opposite is the case. The major implication of this findings for the present study is that it further reduces the likelihood that the bar examination itself is biased. In other words, what differences in performance that are observed between Anglo and minority candidates appear to be primarily due to differences in ability rather than some inherent bias in the test as a whole.

Did Background or Attitude Factors Influence Performance?

A variety of background and attitudinal factors were examined in order to determine whether they may have unduly influenced the performance of minority group members. This issue was examined in two ways. First, the

correlation between each factor and bar performance was computed for each ethnic group. A comparison of these correlations revealed no major significant differences in the degree (strength) of relationship between each factor and Bar performance. For example, the socioeconomic status (SES) of a candidate (as measured by parents' education) was positively related to how well that candidate performed on the Bar. Since minority candidates generally had lower SES levels than their Anglo colleagues, it would seem that this factor might have contributed to the typically poorer Bar performance of minority students. However, this did not seem to be the case in that minority group members still performed far worse than their Anglo colleagues even when their respective parents had the *same* level of education. Thus, the differential in performance is still evident when SES is held constant. Similarly, the relative difference between groups in Bar performance was maintained among those candidates who felt that the exam was *not* a fair test of their legal knowledge or ability to analyze legal problems. In other words, even though proportionately more minority group members adopted this viewpoint, when it was held constant across groups, it did not differentially influence minority group performance on the Bar.

The second method of examining this issue involved determining whether the addition of any one or a combination of background and attitude would significantly improve prediction of Bar performance over that already obtained by using the team of LSAT and LGPA. The logic underlying this approach was that if the addition of such factors improved prediction, then there is reason to believe that such factors may have had a unique influence on Bar performance; such an improvement would be particularly informative if it did not occur equally across ethnic groups. An inspection of the results of this analysis again revealed no unique significant differences between groups nor any substantial improvement across groups in the level of prediction achieved (i.e., over that already obtained by just using the team of LSAT and LGPA).

GRADING PRACTICES STUDY

This Study was undertaken in order to determine whether the cultural background of a grader is related to the scores that grader assigns to essay answers written by candidates whose cultural background was the same as opposed to different from that of the gradee's.

The Study was done in conjunction with the July, 1974 Bar Examination. Four large California law schools assisted in this study by each identifying groups of their Black, Chicano and Anglo graduates who were taking the July, 1974 Bar Examination. An approximately equal number of candidates from each group and from each school was then randomly selected and copies of their answers to the first two essay questions were received from the Committee of Bar Examiners for use in the Study. Three Anglos, three Blacks and two Chicanos served as graders for this Study. These individuals were selected on the basis that they recently passed the Bar and otherwise represented the kind of individuals who would normally be selected by the Committee of Bar Examiners as graders of the essay answers. The answers were typed without editing upon receipt from the Committee of Bar Examin-

ers. This did not influence the outcome of the study as there were no signficiant differences between ethnic groups in the overall quality of their handwriting. A Xerox copy of the typed answers were sent to the graders and they received instructions on the grading of the answers from Mr. McDermott. These instructions were similar to the instructions given to the Bar graders of Questions 1 and 2 during their meeting with the Committee of Bar Examiners on September 6 and 7, 1974 as Mr. McDermott was present at that meeting and tape recorded all of the discussions on these two questions. The Committee of Bar Examiners also supplied the Study with the grades assigned each answer by their graders.

The Study concerned itself with two specific inquiries, first whether graders of different cultural groups agreed on the relative quality of the answers and secondly whether they agreed on the absolute quality of the answers. In order to determine the answer to the first question three different analyses were performed. The first involved assessing the degree to which a grader with a given cultural background tended to agree more with other graders of that background than with graders from a different cultural background. Second the reliability of the grades assigned within each cultural group was computed and finally correlations were computed between the average score assigned to an answer by one group of graders with the average score assigned by another group of graders. On the basis of these analyses, it may be concluded that there was generally good, but far from perfect, agreement regarding the relative quality of the answers.

Although the different groups of graders generally agreed with one another regarding the relative quality of the answers, there was a significant amount of variation on the grades assigned by graders of different cultural groups. Specifically, minority group graders tended to be more lenient than Anglo graders, especially on Question #1, which was considered by the Committee of Bar Examiners to be the more difficult of the two questions. However all three kinds of graders in the Study, as well as the Bar graders agreed that Anglos generally wrote better answers to both Questions #1 and #2, than did minority candidates.

SUMMARY AND CONCLUSIONS

The results of the Item Bias and the Test Bias Studies indicated that there was no practically significant cultural bias in the essay or MBE items, in the MBE subtests, or in the MBE or essay tests as a whole. Further, differences in Bar performance between minority and Anglo candidates were not uniquely attributable to background or attitude factors in the sense that when these factors were held constant, the differences in performance levels between ethnic groups remained. The only significant result of the first study that might warrant a change in current practice was the finding that there were systematic differences between groups in which essay questions they chose to answer. Specifically, Anglos tended to chose questions dealing with elective courses disproportionately more often then did minority candidates.

The results of the Grading Practices Study indicated that there was no statistical or practical evidence for possible grader bias in the scoring of the

essay answers. In other words, minority graders tended to agree with Anglo graders in which answers were considered good or bad irrespective of the ethnicity of the candidates who wrote those answers.

The best predictor of Bar performance were the team of LSAT and LGPA. Knowledge of a candidate's socioeconomic background, attitudes, or ethnicity did not improve the level of prediction achieved by just the combination of LSAT and LGPA. Thus, the difference in Bar performance that exists between Anglo and minority candidates is most likely a function of differences in their respective skills and knowledge rather than in any bias in the Bar examination or in how it is graded.

PART 3

APPENDICES