

AN INVESTIGATION OF POSSIBLE ITEM AND GRADER BIASES
IN A STATE BAR EXAMINATION.¹

Stephen P. Klein²
The Rand Corporation
University of California, Los Angeles

Purpose

Black and Mexican-American candidates generally score substantially lower on state bar examinations than do Anglo candidates. The purpose of the research presented in this report was to determine whether this differential in performance was a function of the relative abilities of the examinees versus biases in the examination's questions and/or the scoring of the answers to these questions.

State Bar Examination

The state bar examination used in this research had two major sections. One section, called the "Multistate Bar Examination" (MBE), was composed of 200 multiple choice items. These items were divided equally among five content areas corresponding to basic fields of law, such as torts and contracts. The multistate examination is used throughout the United States and is published and scored by the Educational Testing Service.

The other section of the bar examination was developed by the state's own committee of bar examiners and contained a set of 20 essay questions. Each examinee was required to answer 15 of these 20 questions at the average rate of 52 minutes per question.

Candidates

A sample of 343 Anglos, 128 Blacks, and 83 Chicanos was drawn from among those candidates who in July of 1974 were taking the examination for the first time. Information regarding ethnicity of these candidates was obtained from them via a mailed questionnaire. Data on the candidates' performance in law school and on the Law Schools Admissions Test (LSAT) was obtained from their law school records.

1. Paper presented at the meetings of the Western Psychological Association, Los Angeles, California; April 9, 1976.
2. This study was conducted and prepared while Dr. Klein was a private consultant to the State Bar Association of California. The views expressed in this report are his own and do not necessarily represent those of the State Bar Association, The Rand Corporation, or the University of California.

Analysis of Bias Across MBE Subtests

The first analysis of the bar examination investigated whether the relative differences between ethnic groups was essentially the same across the five content areas in the Multistate section of the test. In other words, were certain subtests unusually more or less difficult for certain groups? This question was examined by means of a repeated measures analysis of variance. In this context, "bias" was defined as a significant interaction between ethnic group and subtests. The results of this analysis appear in Tables 1 and 2. An inspection of these data indicates a highly statistically significant but practically small percentage of systematic variance accounted for by the interaction effect.

Analysis of Bias Within MBE Subtests

The second analysis examined whether the relative difference between ethnic groups was essentially the same across all 40 items within each subtest of the Multistate examination. This issue was investigated by a series of repeated measures analyses of variance; i.e., one analysis for each of the five subtests. The results of these analyses appear in Table 3. Inspection of these data indicates a few interactions between ethnic group and items that achieve statistical significance, however, none of these effects account for any substantial portion of the variance in candidates' performance.

Analysis of Bias Across Essay Questions

The third analysis investigated whether the differences in performance on the essay section between ethnic groups remained relatively constant or whether certain questions were atypically hard or easy for a given group. A repeated measures analysis of variance was employed with the candidates' scores on the 18 questions that were answered most frequently. The results of this analysis appear in Table 4. These results were consistent with those noted previously in that a very small but statistically significant interaction was obtained.

The foregoing set of studies indicated that a total of less than five percent of the variation in a candidate's score could be attributable to any biases in the kinds of questions asked on the multistate and essay portion of the bar examination. For the reader's information, Tables 5 and 6 contain summary information regarding the candidates' performance on this examination for those individuals for whom complete data were available.

Analysis of Grading Practices

The present research also investigated one more possible source of bias in the examination process; namely, whether there was any bias in the scoring of the answers to the essay questions as a result of the state generally using Anglo graders for this function. In other words, would the difference in performance between groups on this portion of the test diminish if there were greater use of Black and Chicano graders? For reasons that have to do more with the politics rather than the science of research, it was not possible to obtain the essay answers written by the candidates employed in the analyses described previously. Thus, in order to investigate the issue of grader bias, it was necessary to obtain answers written by Anglo, Black, and Chicano candidates on the 1975 examination.

The answers to two questions were selected for this study. Each candidate's answer was then scored independently by three Anglo, three Black, and two Chicano graders. Except for ethnicity, these graders were comparable to those used by the state in terms of both demonstrated satisfactory performance on the bar and training in how to score the written answers. The score assigned to each answer by the state bar's grader also was available as were data on the candidates' performance on relevant predictor measures.

The median intraclass correlation among graders within each group of graders was .77 for question #1 and .76 for question #2. In general, Anglo graders tended to agree with one another slightly more than did the graders within the other two groups.

The results of a repeated measures analysis of variance appear in Table 7. An inspection of these data indicate that no statistically significant interaction was obtained between types of candidates and graders. In other words, the ethnicity of the grader did not differentially influence the scores that grader assigned to answers written by candidates in the various ethnic groups. In fact, the only significant difference between types of graders was that minority group graders tended to be somewhat more lenient than Anglo graders. This finding is illustrated in the column means in Tables 8 and 9. Table 10 presents the correlations between the average score assigned by each group of graders with other variables of interest, such as the number of words written on each answer.

Summary and Conclusions

The analyses presented in this report indicate that there is little evidence of ethnic bias in either the questions asked or in who scores them. Thus, the observed differences in average performance between ethnic groups appears to be primarily a function of their respective abilities to take the bar examination rather than any inherent bias in the test itself. Possible reasons for this difference are beyond the scope of this paper but are likely to involve factors that influence a candidate even before he or she ever enters law school. Finally, it should be noted that this study only examined certain sources of bias within the test itself. It did not examine whether performance on the test was in any way related to being an effective lawyer. While we are waiting for the jury to give us a verdict on that issue, it seems reasonable to continue with the present examination process in that it does not appear to be unduly biased for or against any ethnic group.

TABLE 1

Summary of repeated measures ANOVA across MBE subtests.

Source	df	MS	F	p
Ethnic Group	2	2237.02	35.22	.001
Error	551	46.77		
Subtests	4	4351.16	492.64	.001
Group X Subtests	8	39.05	4.42	.001
Error	2204	8.83		

TABLE 2

Mean MultiState subtest scores.

	Anglos (N=342)	Blacks (N=128)	Chicanos (N=82)	Subtest Average
Real Property	22.19	18.34	18.49	20.74
Contracts	28.38	26.60	26.72	27.72
Torts	29.34	26.57	27.24	28.39
Criminal	27.12	24.59	25.18	26.24
Evidence	28.72	25.79	26.40	27.70
Group Average	27.15	24.38	24.81	26.16

TABLE 3

Summary of repeated measures ANOVA for each MBE subtest.

Source	df	Real Property		Contracts		Torts		Criminal		Evidence	
		MS	F	MS	F	MS	F	MS	F	MS	F
Ethnic Group	2	23.45	43.58**	4.88	14.22**	10.54	31.96**	8.87	21.84**	12.09	27.77**
Error	551	.53		.34		.33		.41		.44	
Items	39	114415.56	33.84**	266965.31	101.58**	134343.38	85.94**	87614.31	34.55**	846584.19	405.54**
Group X Items	78	6348.92	1.87**	1619.28	.62	960.62	.61	5371.28	2.12**	5488.82	2.63**
Error	21489	3380.20		2628.06		1563.21		2536.24		2087.55	

** $\alpha \leq .01$

TABLE 4

Summary of ANOVA for Essay Questions

Source	df	MS	F	P
Ethnic Group	2	24605.91	35.22	.001
Error	551	698.64		
Items	17	1371.04	20.03	.001
Ethnic Group X Items	34	139.70	2.04	.001
Error	9367	68.44		

TABLE 5

Mean score on each measure for each group.

Measure	Anglos (N=248)	Blacks (N=82)	Chicanos (N=52)
UGPA	2.94	2.70	2.67
LSAT	567.39	443.08	461.94
LGPA	2.74	2.22	2.25
MBE	478.45	428.02	432.50
Essay	1135.94	1041.03	1039.90
Total	1614.40	1469.05	1472.40

TABLE 6

Correlations between predictors and criterion variables for each group.

Predictor(s)	Criterion	Anglos	Blacks	Chicanos
LSAT	MBE	.51	.59	.22
	Essay	.36	.37	.24
	Total	.46	.51	.24
UGPA	MBE	.12	.04	.01
	Essay	.17	.03	.14
	Total	.16	.04	.09
LGPA	MBE	.39	.31	.47
	Essay	.40	.29	.51
	Total	.44	.33	.53
LSAT & LGPA	MBE	.62	.59	.48
	Essay	.55	.42	.55
	Total	.56	.54	.55

TABLE 7
 Summary of results for repeated measures ANOVA
 for Questions #1 and #2

Source	df	Question #1			Question #2		
		MS	F	p	MS	F	p
Candidates	2	2210.94	11.02	.001	937.70	5.34	.007
Error	76	200.62			175.50		
Graders	3	1055.06	37.40	.001	596.91	23.19	.001
Candidates X Graders	6	41.67	1.47	.187	42.14	1.64	.138
Error	228	28.21			25.79		

TABLE 8
 Cell and marginal means for Question #1.

Type of Candidate	State Bar	Type of Grader			Row Mean
		Anglos	Blacks	Chicanos	
Anglo (N = 28)	70.71	70.65	72.69	77.66	72.93
Black (N = 25)	60.40	59.67	66.87	68.90	63.96
Chicano (N = 26)	65.58	63.53	68.21	71.25	67.14
Column Mean	65.76	64.83	69.37	72.78	68.18

TABLE 9
 Cell and marginal means for Question #2

Type of Candidate	State Bar	Type of Grader			Row Mean
		Anglos	Blacks	Chicanos	
Anglo (N = 28)	74.29	71.90	72.98	76.77	73.98
Black (N = 24)	70.63	64.58	70.35	71.77	69.33
Chicano (N = 27)	68.33	64.06	70.49	71.85	68.57
Column Mean	71.14	67.00	71.18	73.57	70.72

TABLE 10

Correlations between the average scores assigned by each type of grader and related factors. ^{1,2}

	Anglo Graders	Black Graders	Chicano Graders	State Bar Grader	LSAT Score	LGPA	Total Bar Score	No. words written on question #1
Anglo Graders	(.43)**	.79**	.75**	.69**	.42**	.55**	.58**	.55**
Black Graders	.60**	(.32)**	.67**	.53**	.39**	.59**	.44**	.51**
Chicano Graders	.67**	.61**	(.16)	.69**	.32**	.49**	.47**	.60**
State Bar Grader	.67**	.59**	.66**	(.24)*	.28**	.42**	.44**	.48**
LSAT Score	.30**	.30**	.29**	.32**	---	.52**	.59**	.22*
LGPA	.42**	.37**	.24*	.38**	.52**	---	.72**	.34**
Total Bar Score	.57**	.56**	.47**	.62**	.59**	.72**	---	.34**
No. words written on question #2	.44**	.31**	.26*	.22*	-.03	.14	.22*	(.29)**

* $\alpha \leq .05$ ** $\alpha \leq .01$

1. Correlations for Questions #1 and #2 appear above and below the main diagonal, respectively.
2. Correlations between Questions #1 and #2 appear along the main diagonal in parentheses and are based on those candidates (N = 69) who answered both questions.