

81-1-12 PK

FACTORS ASSOCIATED WITH THE DIFFERENCE IN PASSING RATE BETWEEN
ANGLO AND HISPANIC APPLICANTS ON THE NEW MEXICO BAR EXAMINATION.

Stephen P. Klein, Ph.D.
January 12, 1981

INTRODUCTION

An analysis of the background characteristics of Anglo and Hispanic applicants taking the New Mexico Bar Examination (NMBE) indicates that Anglo applicants tend to have higher law school grades and admission test scores. And, a higher percentage of Anglo than Hispanic applicants pass the NMBE. The research described in this report was undertaken to determine the extent to which the difference in passing rates between these groups can be explained by differences between them in their respective law school grades and admission scores versus explained by factors that are related solely to group affiliation. In other words, if the difference between groups on the NMBE is wider than what would be expected on the basis of performance level differences between them prior to taking the examination, then the NMBE can be said to be biased against Hispanic applicants.

MEASURES

The following data were available for graduates of the University of New Mexico's School of Law who took one or more of the ten NMBE's that were administered between February, 1976 and July, 1980:

- o Anglo/Hispanic Group Affiliation (AHGA). Applicants taking the NMBE were required to provide a set of their fingerprints. And, it is standard law enforcement procedure to indicate group affiliation on the card used for recording fingerprints. Thus, obtaining group affiliation data on the applicants was a serendipitous by-product of the fingerprint requirement.
- o Law School Admission Test (LSAT) score. The LSAT is administered by the Educational Testing Service of Princeton, New Jersey. Almost all American Bar Association approved law schools use the LSAT as part of their admissions procedures because scores on this test generally contribute to the prediction of law school grades.
- o Law School Grade Point Average (LGPA). The University of New Mexico indicated LGPA on a four point scale with a 4.0 being equivalent to an A, 3.0 to a B, etc.
- o Multistate Bar Examination (MBE) score. The MBE is developed by the National Conference of Bar Examiners and administered by the Educational Testing Service. It is used as part of the bar examination in about 47 jurisdictions. The test consists of 200 multiple choice questions drawn from the following six content areas: constitutional law, contracts, criminal law, evidence, real property, and torts. MBE scores are scaled (i.e., adjusted) by ETS so as to control for possible differences in test difficulty across administrations.

- o New Mexico Essay Examination (Essay) score. The Essay portion of the NMBE consists of 24 essay questions. Approximately 40 minutes is given to answer each question, but applicants are allowed some flexibility in how they allocate their time. Most of the questions are developed by New Mexico's Board of Bar Examiners. The Board draws the remaining questions from the bank of questions maintained by the National Conference of Bar Examiners. All the answers to a given question are graded under the direction of the Board by a single attorney.

PRELIMINARY ANALYSES

Tables 1 and 2 contain statistical data for all the applicants who took one or more of the ten examinations. An inspection of these tables indicates that the applicants taking the examination in July tended to do better on the MBE than those taking it February. This is not surprising in that a larger proportion of the February takers than July takers were repeaters (i.e., applicants who failed the NMBE previously). The absence of a comparable difference in average essay scores between July and February takers coupled with the relatively strong correlation between MBE and Essay scores suggests that the essay portion of the examination tended to have easier questions and/or be graded more leniently in February than in July. Variations in essay test difficulty also are apparent within the February and July examinations. For example, although the average MBE scores on the July, 1977 and 1980 examinations were essentially the same, the average score on an essay question was almost six points lower on the 1980 examination.

The foregoing differences in essay grading standards precluded the pooling of essay scores across administrations; e.g., an essay score of 70 on one NMBE might not denote the same level of performance as a score of 70 on another NMBE. And, such pooling is necessary in order to have a sufficiently large number of applicants on which to base conclusions. Since the MBE scores were already adjusted for possible differences in test difficulty across administrations and since MBE and Essay scores were highly correlated with one another, it was decided to correct the problem of variations in essay test difficulty by scaling the essay scores to the same mean and standard deviation as the MBE scores. This scaling was done for each of the ten examinations. In accordance with New Mexico's past policies, each applicant's Total score was computed using the following formula: $Total = (.667)(Essay) + (.333)(MBE)$.

All the applicants taking a given NMBE were used for scaling that NMBE's essay scores. All of the other analyses and tables presented in this report are limited to University of New Mexico (UNM) graduates. This restriction was imposed because of the importance of LGPA to the study and the lack of LGPAs (or the ability to equate them to UNM's standards) for non-UNM graduates. Moreover, if an applicant took the NMBE more than once, only the data from the first testing were used in these subsequent analyses. This was done so that no applicant would have a disproportionately large effect on the results.

Table 1

MEANS, STANDARD DEVIATIONS, AND OTHER SUMMARY STATISTICS ON THE FIVE FEBRUARY EXAMINATIONS GIVEN BETWEEN 1976 AND 1980.

Summary Statistical Data	1976	1977	1978	1979	1980	Mean
MBE - Mean Scale Score	137.4	140.2	137.0	135.1	135.1	137.0
Standard Deviation	15.6	14.6	14.9	15.6	18.9	15.9
Essay - Mean Question Score	70.3	69.7	68.0	66.1	63.0	67.4
Standard Deviation	5.4	5.8	5.9	7.6	8.7	6.7
Essay Test Reliability	.78	.81	.79	.86	.87	.83
MBE/Essay Correlation	.77	.72	.66	.66	.82	.73
Percent Passing	76	74	61	71	59	68
Score Needed for Passing*	127	131	133	127	131	130
Number of Applicants	89	112	104	107	95	101
Percent Anglo	73	79	61	64	67	68.8
Percent Hispanic	26	18	36	33	22	27.0

Table 2

MEANS, STANDARD DEVIATIONS, AND OTHER SUMMARY STATISTICS ON THE FIVE JULY EXAMINATIONS GIVEN BETWEEN 1976 AND 1980.

Summary Statistical Data	1976	1977	1978	1979	1980	Mean
MBE - Mean Scale Score	146.4	140.9	142.7	141.7	141.2	142.6
Standard Deviation	16.6	15.9	17.5	16.1	15.4	16.3
Essay - Mean Question Score	68.0	70.0	65.5	70.9	64.2	67.7
Standard Deviation	5.5	7.5	6.3	6.6	7.3	6.6
Essay Test Reliability	.78	.87	.85	.85	.82	.84
MBE/Essay Correlation	.77	.77	.76	.75	.77	.76
Percent Passing	66	68	67	84	65	70
Score Needed for Passing*	144	136	136	126	134	135
Number of Applicants	137	221	197	183	196	186
Percent Anglo	80	75	71	73	82	76.2
Percent Hispanic	19	23	26	22	14	20.8

*Essay scores were scaled to the same mean and standard deviation as the MBE scores. A Total score was computed for each applicant using the following formula: $(.667)(\text{Essay}) + (.333)(\text{MBE})$. The "score needed for passing" is the Total score that would have passed the same percent of applicants as actually did pass; e.g., 66 percent of the applicants on the July, 1976 examination had Total scores of 144 or higher.

ANALYSIS PLAN

If the NMBE is biased against Hispanic applicants, then the gap in average performance level between Anglo and Hispanic applicants on the NMBE would be greater than the differences between these two groups in their average LSAT scores and LGPAs. If this occurred, then knowledge of an applicant's group would allow one to make a more accurate prediction of that applicant's NMBE scores than could be made solely on the basis of that applicant's law school grades and admission scores. For example, if the NMBE tended to widen the gap, then one would predict that an Anglo applicant would do better on the NMBE than a Hispanic applicant even if they had comparable LSAT scores and LGPAs. Thus, bias can be measured by assessing the degree to which knowledge of an applicant's group improves the prediction of NMBE scores over the level of predictive accuracy achieved by the use of LSAT scores and/or LGPAs alone.

A standard statistical technique that can be used for measuring predictive accuracy is called regression analysis. This procedure provides an index of the degree to which score differences between applicants on one variable (such as the MBE) can be explained by (i.e., correspond with) score differences between these same applicants on some other variable (such as LSAT). If no systematic relationship exists between two variables, then the percent of variance explained is zero. If knowledge of an applicant's score on one variable permits making a perfect prediction of that applicant's score on another variable, then 100% of the variance in one variable is explained by the variance in the other. For example, if the applicants were rank ordered on the LSAT in exactly the same way as they were ordered on the MBE (i.e., the applicant with the highest LSAT also had the highest MBE, the applicant with the second highest LSAT had the second highest MBE, etc.), then 100% of the variance in MBE scores would be explained by the variance in LSAT scores.

Regression analysis also can be used to determine how well a combination of measures (such as LSAT + LGPA or LSAT + LGPA + AHGA) predicts scores on some other variable (such as the MBE). However, because predictors may share explanatory power, a combination of measures may account for less variance than the sum of their separate abilities to predict. For example, although LGPA and LSAT scores may each explain about 45 percent of the variance in MBE scores, the combination of LGPA and LSAT may only explain a total of 50 percent because they share explanatory power. This sharing comes about as a result of the correlation between LGPA and LSAT; i.e., applicants who tend to get high LSAT scores also tend to get high LGPAs.

In summary, the NMBE would be biased if the amount of variance in NMBE scores that is explained by LSAT and LGPA could be increased by consideration of an applicant's group. The degree of bias is measured by the amount of increase...the bigger the increase, the greater the bias. The practical consequence of any bias, however, is indicated by its effect on the passing rate. In other words, in order for any bias that is detected to have policy relevance, it must explain a significant portion of the difference in passing rates between groups.

RESULTS

Table 3 presents the percent of variance in MBE, Essay, and Total NMBE scores that is explained by each predictor singly and in combination with the other predictors. An inspection of these data indicated that:

- o LGPA provided a more accurate estimate of an applicant's likelihood of success on the NMBE than did LSAT which in turn provided a more accurate estimate than did knowledge of the applicant's AHGA.
- o The differences in the relative accuracies of the predictors was most apparent on the MBE portion of the examination.
- o More of the variance in MBE scores than in Essay scores could be explained by all three predictors. Most, if not all, of this difference is due to the MBE being more reliable than the Essay.
- o Adding AHGA to a prediction system that already contained LGPA and/or LSAT usually increased explanatory power by less than three percent.
- o When LSAT scores were teamed with LGPA, predictive accuracy generally improved and the effect of AHGA all but disappeared. For example, the combination of LGPA and LSAT scores explained 57 percent of the variance in MBE scores while LGPA and AHGA explained only 51 percent. LGPA alone explained just 48 percent of the MBE score variance.

Table 3

PERCENT OF VARIANCE IN NMBE SCORES THAT WAS EXPLAINED WHEN PREDICTORS WERE USED SINGLY AND IN COMBINATION WITH OTHER PREDICTORS.

Variables Used to Predict NMBE Scores	Percent of Variance Explained		
	MBE	Essay	Total
Law School Grade Point Average (LGPA)	48	47	54
Law School Admission Test Score (LSAT)	36	18	27
Anglo/Hispanic Group Affiliation (AHGA)	20	13	18
LGPA + AHGA	51	47	56
LSAT + AHGA	38	21	30
LGPA + LSAT	57	48	58
LGPA + LSAT + AHGA	57	48	59

DISCUSSION

The largest incremental effect of AHGA (three percent) was obtained in the prediction of MBE scores when the only other predictor was LGPA. An inspection of the summary statistics for each group that is presented in Table 4 indicates that this increase corresponded to a slightly larger gap in performance level between Anglo and Hispanic applicants on the MBE (1.1 standard deviation units) than on LGPA (.9 standard deviation units).

Table 4

MEANS, STANDARD DEVIATIONS, AND OTHER SUMMARY STATISTICS FOR THE ANGLO, HISPANIC, AND TOTAL SAMPLES.

Variable		Total Sample	Group		Difference Between Groups*
			Anglo	Hispanic	
LGPA	Mean	2.81	2.90	2.51	0.9
	SD	.44	.42	.38	
LSAT	Mean	594.4	619.9	503.9	1.6
	SD	85.7	70.2	73.9	
MBE	Mean	143.9	147.6	130.9	1.1
	SD	14.8	13.1	13.1	
Essay	Mean	141.7	144.4	131.9	1.0
	SD	13.7	12.6	12.8	
Total	Mean	142.4	145.5	131.5	1.1
	SD	13.1	11.8	11.5	
Percent Passing		78	86	52	34
Number of Applicants		441	344	97	247

*The differences in mean scores between groups are presented in terms of standard deviation units. The standard deviation in the total sample was used for this purpose.

The addition of AHGA to LGPA increased the amount of explained variance in MBE scores by three percent whereas the addition of LSAT increased it by nine percent (see Table 3). The inclusion of possible interactions between the predictors failed to increase the amount of explained variance. Thus, AT MOST, only three percent of the variance in MBE scores is due to the unique effect of group affiliation. The rest is attributable to the unique contribution of LSAT (i.e., that part of LSAT's explanatory power that is not shared by either LGPA or AHGA); LGPA itself; the less than perfect reliability of LGPA, LSAT, and NMBE scores; and other factors not examined in this research.

The addition of AHGA to LGPA did not improve prediction of Essay scores. The addition of AHGA to LSAT increased the amount of variance explained by LSAT alone by three percent (i.e., from 18 to 21 percent). However, the combination of LSAT and AHGA explained less than one half of that accounted for by LGPA alone. Moreover, the combined unique contribution of LSAT and AHGA (i.e., after LGPA was already in the prediction system) was only one percent. It is apparent, therefore, that the differences between groups on the Essay portion of the examination are not due to any systematic bias that is related to group affiliation.

The pattern of results with Total NMBE score essentially reflected those obtained with the MBE and Essay components; i.e., group affiliation explains less than three percent of the variance in bar scores. It is important to note, however, that the variance explained in Total score (59 percent) is slightly more than that explained in MBE scores (57 percent) and substantially more than that accounted for in Essay scores (48 percent). These differences are largely attributable to the respective reliabilities of the three NMBE scores (.90 for MBE, .84 for Essay, and .92 for Total). In other words, it is more difficult to predict Essay scores because they are more affected by chance factors, such as those that are introduced by the subjective nature of essay grading procedures. In short, the Total score provides a more accurate indicator of an applicant's performance level with respect to the skills and knowledge measured by the NMBE than does either the MBE or Essay portions alone.

The unique contribution of AHGA to the prediction of NMBE scores can be translated into the number of points an Anglo applicant would, on the average, be expected to score higher than an Hispanic applicant if these two applicants had the same set of LGPA and LSAT scores. The results of this analysis indicated that the Anglo applicant would be expected to score 2.4 points higher on the MBE, 1.9 points higher on the Essay, and 2.1 points higher on Total score.* For example, if an Anglo applicant had an LSAT score of 600 and an LGPA of 2.85, then that applicant would be expected to have a Total NMBE score of 143. An Hispanic applicant with an LSAT of 600 and an LGPA of 2.85 would be expected to obtain a Total score of 141.

If the passing score was set at the usual February level of 130 and if two points were added to every Hispanic applicant's Total NMBE score, then the Hispanic pass rate would increase from 56 to 60 percent. If the passing score was set at the usual July level of 135, then adding two points to every Hispanic's score would increase the Hispanic passing rate from 38 to 42 percent. In short, correcting for the unique effect of AHGA on bar scores resulted in increasing the Hispanic pass rate by about four percent.

* Statistical Note: A multiple regression equation was constructed for each NMBE score. The predictors in this equation and the order in which they were entered were as follows: LGPA, LSAT, and AHGA. The latter variable was constructed by assigning Hispanics a value of zero and Anglos a value of one. Thus, the unstandardized regression weight for AHGA indicated the number of points on the dependent variable that would be affected by the inclusion of AHGA. For example, the equation for predicting Total NMBE score was:

$$\text{Total} = (18.6828)(\text{LGPA}) + (.02966)(\text{LSAT}) + (2.0661)(\text{AHGA}) + 69.81$$

The preceding analyses were repeated with UNM graduates who took the test for the first time in July. The results of this study were, with two exceptions, the same as those described above. One of these exceptions was that the amount of variance explained increased by two percent. The other exception was that AHGA's contribution to the total variance explained was cut in half. Some of the factors that may have produced these differences in the two sets of analyses are: (1) the July UNM first timers may have been qualitatively different than the February first timers in terms of courses taken, professors, etc...they certainly did not follow the typical pattern of completing law school in May and taking the NMBE two months later; (2) more sampling error may have been introduced by scaling the Essay scores in February then in July because of the fewer number of applicants on which to base the February scaling; and (3) including February takers doubled the number of tests but increased the number of applicants by only 15 percent which in turn may have reduced the comparability of scores across applicants. In short, the increase in the amount of variance explained and the reduction in the unique contribution of AHGA by limiting the sample to July takers probably came from decreasing the effect of extraneous factors.

A similiar study, conducted on one administration of the California bar examination, also found a large difference between Anglo and Hispanic passing rates. However, AHGA did not increase the percent of explained variance in MBE and/or California essay scores over that explained by differences between groups in their law school grades and LSAT scores.

An analysis of the relationship between NMBE scores and an applicant's sex group indicated that knowledge of an applicant's sex could not be used to increase the amount of variance explained over that already explained by LGPA and/or LSAT. It was noted, however, that on the average, female applicants tended to receive higher LGPAs than male applicants. Women also tended to slightly out perform men on LSAT and on the bar examination.

Finally, an inspection of the data in Table 4 indicated that the size of the gap in performance between Anglos and Hispanics was somewhat greater on LSAT than on LGPA. It was further observed that LSAT explained more of the variance in Anglo LGPAs (12 percent) than in Hispanic LGPAs (8 percent). In the total sample, 19 percent of the variance in LGPA corresponded with the variance in LSAT. These findings suggest that LSAT is not as good a predictor of Hispanic performance at UNM's law school as it is of Anglo performance. Some, but certainly not all, of the factors that could produce this pattern of results are: (1) LSAT underestimating the true performance level of Hispanic applicants in law school (such as might stem from differences in motivation, time spent studying, test-wiseness, and attitudes toward multiple choice tests), (2) differences in the difficulty of the courses taken by the two groups, (3) professors being reluctant to give grades below C even though there may be a difference in the average performance level of Anglo and Hispanic students who received C's (i.e., the fact that the average grade of UNM graduates is a B suggests that there might be an artificial "cellar" on LGPA and Hispanic applicants may be more likely to fall closer to the bottom of that cellar than Anglos), and (4) the lower reliability of LGPA (.85) as compared to LSAT (.90) and the less than perfect reliability of both of them. Some of these same factors as well as others also may have contributed to the small percent of variance in NMBE that was uniquely attributable to group affiliation.

SUMMARY AND CONCLUSIONS

This study investigated whether New Mexico's bar examination exacerbated or merely mirrored the differences between Anglo and Hispanic applicants that are present in their average law school grades and admissions scores. The applicants used in this analysis were graduates of the University of New Mexico's School of Law who took the bar examination for the first time between February, 1976 and July, 1980.

The analytic procedures employed in this research involved determining the degree to which knowledge of an applicant's group (i.e., Anglo versus Hispanic) permitted making a more accurate prediction of that applicant's bar scores than could be made solely on the basis of the applicant's law school grade point average (LGPA) and admissions test (LSAT) scores.

The initial results of these analyses indicated that slightly more of the differences between applicants in their bar scores could be predicted from knowledge of an applicant's group than could predicted solely on the basis of an applicant's LGPA and LSAT score. If bonus points were awarded to Hispanic applicants to make up for this difference, then about four percent more Hispanic applicants would pass the examination than would otherwise be the case.

When the sample was restricted to University of New Mexico graduates who took the test for the first time in July, a slightly more accurate prediction of bar scores was obtained; and, even less of the gap between Anglos and Hispanics on the examination could be explained by group affiliation. Moreover, a study conducted on the MBE and Essay portions of California's bar examination found that group affiliation did not provide any unique contribution to the prediction of bar scores. These two sets of findings suggest that some of the variance in New Mexico's examination that was uniquely associated with group affiliation in the combined sample (i.e., July plus February) probably came from extraneous factors, such as the particular samples used. In other words, if any bias exists in the New Mexico examination, it probably does not affect the Hispanic passing rate by more than two percent.

It may be concluded, therefore, that at most, only a very small amount of the difference in passing rates between Anglo and Hispanic applicants (i.e., zero to four percent) is likely to be attributable to any bias in New Mexico's bar examination. In contrast, a substantial portion of the difference between groups in their passing rates is associated with differences in their respective law school grades and admissions scores. The remaining portion of the difference in passing rates is due to the unreliability of the measures and to factors that are not correlated with law school grades, admission scores, or group affiliation.

Some of the other findings of this research were as follows: (1) female applicants tended to get higher scores than male applicants on all the variables, (2) the July administration of the examination was harder to pass (as indicated by the level of performance required for passing) than the February administration, (3) there has been variation in the standard of performance required for passing within both the February and July administrations, and (4) LSAT scores under predict the grade point average an Hispanic applicant is likely to receive at UNM's law school. This latter finding, however, may be due as much if not more to problems with law school grades than to shortcomings of the LSAT.

REFERENCES

- Carlson, A. B. and Werts, C. E. Relationships among law school predictors, law school performance, and bar examination results. Princeton, N.J.: Educational Testing Service, 1976.
- Klein, S. P. An investigation of possible item and grader biases in a state bar examination. Paper presented at the meetings of the Western Psychological Association, Los Angeles, California; April 9, 1976.
- Klein, S. P. An analysis of the relationships between bar examination scores and an applicant's law school, admissions test scores, grades, sex, and racial/ethnic group. Paper presented at the meetings of the American Bar Association, Dallas, Texas; August, 1979.

GLOSSARY OF STATISTICAL TERMS

- Variable** Any characteristic of persons on which they differ. Law school grade point average, LSAT, and the MBE and Essay portions of a bar examination are all examples of variables.
- Mean** The mean is the arithmetic average. It is computed by summing all the scores on a variable and then dividing by the number of scores that entered that sum.
- Standard Deviation (SD)** The standard deviation is an index of the degree to which scores on a variable spread out on either side of the mean. The larger the standard deviation, the greater the spread. If the shape of the score distribution is normal, then about 34 percent of the scores fall between the mean and one standard deviation above it. The standard deviation of a set of scores is computed by taking the square root of its variance.
- Variance** The variance of a set of scores is the average of the squared difference between each score and the mean score.
- Percent of Explained Variance** This is an indicator of the extent to which individuals relative standing on one variable (i.e., how much their scores deviate from the mean score on that variable) generally corresponds with their relative standing on some other variable (or set of variables). In other words, the percent of explained variance indicates the extent to which their scores on one variable tends to be systematically related to their scores on some other variable or set of variables. The higher the percent of variance explained (up to a theoretical maximum of 100), the stronger the relationship.
- Regression Analysis** This is a statistical procedure that allows one to measure the percent of variance in one variable that is explained by (i.e., corresponds with, is attributable to, or is systematically associated with) the variance in some other variable or set of variables.
- Scaling** In the context of this report, scaling refers to converting the scores on one variable to the same units of measurement (in terms of mean and standard deviation) as the scores on some other variable.