

Group Differentiated Prediction

Robert L. Linn and C. Nicholas Hastings
University of Illinois at Urbana-Champaign

Studies of predictive bias have frequently shown that a prediction equation based on majority group members tends to overpredict the criterion performance of minority group members. Two statistical artifacts that may cause the overprediction finding are reviewed and evaluated using data for black and white students at 30 law schools. It is shown that (1) the degree of overprediction decreases as the predictive accuracy for white students increases, and (2) that overprediction can be caused by the effects of selection on variables not included in the regression model. Use of Heckman's (1979) procedure to adjust the estimates of the regression parameters was found to essentially eliminate overprediction.

Predictive bias has been the focus of a substantial number of studies in a wide variety of selection situations, including military, employment, and educational settings. The basic paradigm of these studies is quite familiar by now. Within-group regression equations are computed and the standard errors of prediction, the slopes, and the intercepts are compared. If different prediction systems are obtained for two groups, e.g., a minority group and a majority group, or men and women, then the use of an equation based upon one group will result in systematic errors of prediction when applied to the other group. The natural question is then: What is the magnitude and direction of those systematic errors?

A common approach to answering this question is to use the majority group prediction equation to obtain predictions for values of the predictors equal to the minority group means and to compare these predictions to the actual minority group mean on the criterion. Alternatively, predictions based on the two equations may be made for various combinations of the two predictors to define regions where one equation yields higher predictions than the other.

The naive expectation, in keeping with a belief that tests are biased against minority group members, was that the predicted criterion performance from the majority group equation would be lower than the actual performance of minority group members. That is, that there would be a bias against minority group members in the sense that their criterion performance would be underpredicted. However, the results of most studies run counter to this expectation. The bulk of the evidence shows either no difference in the predictions from minority and majority group equations or that the majority group equation tends to overpredict the minority group performance (Linn, 1982; Schmidt & Hunter, 1981). These results led Schmidt and Hunter (1981, p. 1128) to conclude that "cognitive ability tests . . . are fair to minority group applicants in the sense that they do not underestimate expected job performance of minority groups."

Although the equal-prediction or overprediction result is well documented, the explanation of the findings is unclear. The purpose of this paper is to

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 8, No. 2, Spring 1984, pp. 165-172
© Copyright 1984 Applied Psychological Measurement Inc.
0146-6216/84/020165-08\$1.65

review and evaluate some of the bases for the results, giving particular emphasis to statistical artifacts that may lead to the overprediction result. This will be done within the context of prediction of first-year grades in law school for white and black students.

Several previous studies (e.g., Linn, 1975; Powers, 1977; Schrader & Pitcher, 1973, 1974) have investigated the differential prediction of grades in law schools for black and white students. The results of 42 within-school analyses using undergraduate grade-point average (UGPA) and Law School Admission Test (LSAT) scores as predictors were summarized by Linn (1982). The regression systems were found to be significantly different in 38 of the 42 analyses. In no case did the regression equation for whites or the combined-group equation underpredict the actual mean first-year grade average (FYGA) of black students. In 38 of the studies the mean FYGA of black students was overpredicted by the combined-group equation by more than one-fifth of a combined-group standard deviation.

Data Set

The data reported in this paper were obtained from the 1981–82 data files of the Validity Study Service provided by the Law School Admission Service. These files contain predictor scores for applicants as well as predictor and criterion scores for matriculated students. Data from 30 schools that had criterion data for a minimum of 35 black students were used for the analyses described below. The number of black students per school ranged from 36 to 157 with a mean of 60. The *ns* for white students ranged from 48 to 1407 with a mean of 708. In order to protect the identity of the individual law schools, scores on all three variables (i.e., UGPA, LSAT, and FYGA) were transformed to have means of 50 and standard deviations of 10 for the white students within each school.

Within-group regression equations were computed, and the mean predicted score for black students based on the equation for white students was compared to the actual mean FYGA of black students. The frequency distribution of the predicted

minus the actual mean FYGA is reported in Table 1. The results are generally consistent with those reported in 1973 and 1974 by Schrader and Pitcher and those reported in 1977 by Powers. There is a clear tendency for an equation based on white students to overpredict the grades obtained by black students, typically by a little over three-tenths of a standard deviation.

For purposes of this paper, the results in Table 1 are of interest primarily as background. The findings add little to what is already known. As was already indicated, the main question is whether the results may be due, in part or whole, to statistical artifacts. Two factors that may contribute to the overprediction finding will be considered. These are: (1) the precision of the prediction for the majority group within a school, and (2) distortions of the majority group regression equation caused by selection.

Predictive Accuracy

It is well known that errors of measurement in a predictor reduce the slope of a simple regression line. Using classical test theory assumptions, the slope of the regression of a criterion measure on observed predictor scores can be shown to be equal to the slope of the regression of the criterion on the predictor's true scores times the reliability of the predictor. Linn and Werts (1971) used this result to show that errors of measurement can cause overprediction for a minority group when there would be none in the absence of errors of measurement. This phenomenon may be conceptualized as a special case of a more general one that suggests that the less accurate the predictions are for the majority group the greater will be the tendency for the majority-group equation to overpredict the performance of a lower-scoring minority group.

To illustrate this, imagine an idealized predictive composite, C^* , that has a maximum correlation with the criterion, Y . Assume that the actual criterion composite, C , is equal to the idealized composite plus a random error, E ,

$$C = C^* + E \quad (1)$$

The expected value of E is assumed to equal zero within both the minority and majority groups, and

Table 1
Distribution of the Difference Between the Predicted FYGA
Using the Equation for White Students with Predictors
Equal to the Black Student Means and the Actual
Mean FYGA for Black Students at 30
Law Schools*

Difference	Frequency	Number of Black Students Within Each School
11	1	36
10	0	
9	1	44
8	1	157
7	2	36,74
6	1	81
5	3	36,47,126
4	5	36,40,43,94,100
3	4	43,53,65
2	7	38,44,44,48,51,53,85
1	3	51,67,102
0	1	38
-1	1	36
Unweighted Mean	3.6	

*FYGA is scaled to have a standard deviation of 10 for white students within each school.

a common regression of the criterion on C^* is assumed, i.e., there is no predictive bias for C^* . When the standard deviation of Y equals 1, C has a variance equal to the squared multiple correlation (SMC) and the slope of the regression of Y on C is equal to one. Finally, using classical test theory assumptions, the slope of the regression of Y on C^* can be shown to be

$$\text{Slope} = \frac{\text{SMC}}{\text{SMC} - V(E)}, \quad (2)$$

where $V(E)$ is the variance of E . With these assumptions, the overprediction, Res, resulting from the use of C in place of C^* is

$$\text{Res} = \left(\frac{V(E)}{\text{SMC} - V(E)} \right) D, \quad (3)$$

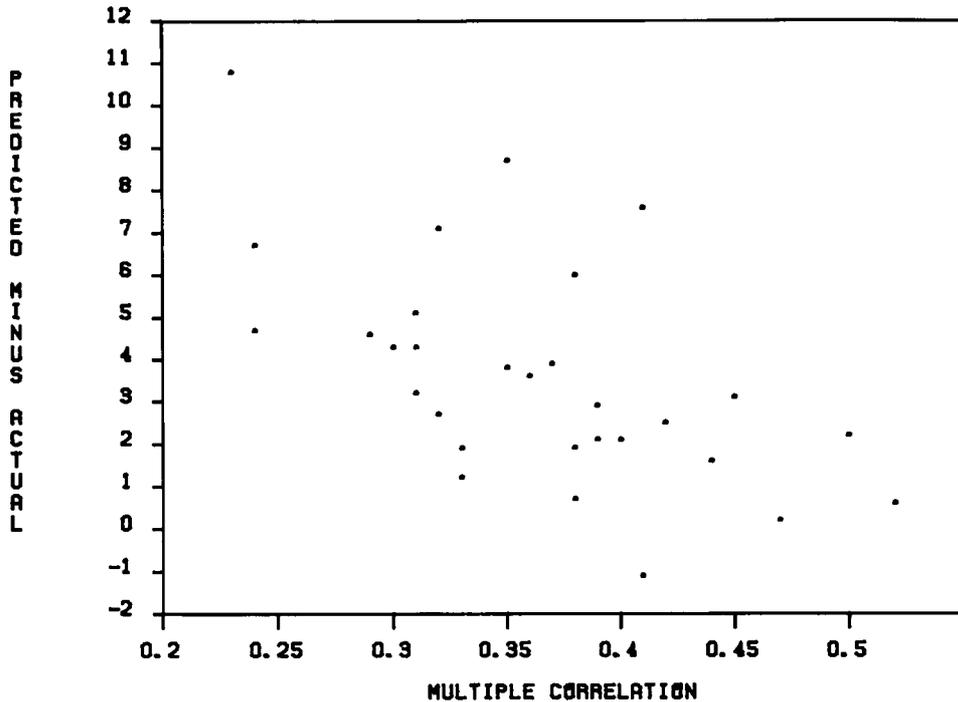
where D is the difference between the majority and minority groups on C .

The above argument is obviously rather contrived and involves a number of possibly unrealistic assumptions. The main point, however, is simple. The amount of overprediction that is expected in-

creases directly as a function of $V(E)$ and inversely as a function of the SMC. Furthermore, $V(E)$ would be expected to decrease as the SMC increases. Thus, the more accurate the prediction for the majority group (i.e., the larger the SMC), the smaller the overprediction that would be expected. This conclusion is also consistent with another argument presented by Linn and Werts (1971) that showed that overprediction can be produced by the failure to include all valid predictors in the regression equation.

The data from the 30 law schools provides some support for the above contention. A plot of the difference between the predicted and observed criterion performance at the 30 law schools against the magnitude of the multiple correlation for majority group students is shown in Figure 1. As would be expected from the above argument, there is a negative relationship between the size of the multiple correlation and the magnitude of the overprediction. The correlation between these two variables is $-.58$.

Figure 1
Scatterplot of Predicted Minus Actual Mean FYGA for Black Students Against the Multiple Correlation for White Students at 30 Law Schools



Selection Effects

The relationship between the degree of predictive accuracy within a school and the amount of overprediction for blacks is of interest. This relationship may itself be due to another statistical artifact, however. It is well known that selection tends to reduce the correlation between a predictor or set of predictors and a criterion measure. Whether or not selection also affects the weights in a regression equation, and thereby influences the degree of overprediction, depends on the rules used for selection. As will be shown, it is quite possible for selection not only to reduce the correlation but to flatten the slope of a regression line and increase the intercept—changes that would produce overprediction for a minority group with lower scores on the predictor (Linn, 1983).

For simplicity, only the single predictor case will be considered. The effect of selection on overpre-

diction shown in this case generalizes to situations with multiple predictors, however. Selection based solely on the predictor will increase the sampling error of the regression coefficients, but such selection does not bias the estimates. Hence, selection might be expected to increase the between-study variation in overprediction or underprediction, but not to produce a systematic tendency in one direction or the other. On the other hand, selection that is not based solely on the predictor of interest results in biased estimates of the regression coefficients as well as increases in their sampling errors. Furthermore, it is seldom the case that selection is based solely on the predictor of interest.

Law schools depend heavily on undergraduate grades and LSAT scores in selecting students. However, as Evans (1977, p. 579) has indicated, law schools also consider such factors as “quality of undergraduate institution, major field, explanations of discrepancies in records, graduate study,

work experience, extracurricular activities, significant achievements, letters of recommendation, residence, and personal statements filed by applicants.” Even if law schools based their decisions solely on UGPA and LSAT, the sample available for analysis would still be subject to other selection effects due to student self-selection and attrition. Students who are accepted by a school but decide to go elsewhere and students who enroll but fail to complete their first year are unavailable for analysis.

The bias in the regression coefficients caused by selection can be seen by considering a simple case. Assume that the regression of Y on X in the unselected population is linear. Then the slope of the regression line is

$$\beta_1 = \rho_{xy} \left(\frac{\sigma_y}{\sigma_x} \right) \quad (4)$$

using the obvious notation. Now assume that a subpopulation is selected on the basis of a third variable, U , which may be highly, but not perfectly correlated with X . Using the standard Pearson-Lawley (Lawley, 1943) assumptions, the slope of the regression of Y on X in the selected population can be shown to be

$$\beta_1^* = \left(\frac{\rho_{xy} - w\rho_{ux}\rho_{uy}}{1 - w\rho_{ux}^2} \right) \left(\frac{\sigma_y}{\sigma_x} \right) \quad (5)$$

where w is one minus the ratio of the variance of U in the selected to the unselected populations.

Not only are the slopes in the selected and unselected populations generally unequal, but the regression in the selected population is nonlinear. As shown by Muthén and Jöreskog (1983), the conditional means of Y given X in the selected population are

$$E^*(Y|X) = \beta_0 + \beta_1 X + \beta_2 f(\lambda) \quad (6)$$

where λ is a linear function of X and $f(\lambda)$ is the ratio of the normal density function to the normal distribution function.

The regression function of a dependent variable, Y , on a covariate, X , for the selected and unselected populations as well as the conditional means in the selected population are illustrated in Figure 2. For this example, it was assumed that selection was based on a third variable, U , that is correlated .89 with X and .56 with Y . The slope of the regression

of Y on X is .6 in the unselected population but only .44 in the selected population.

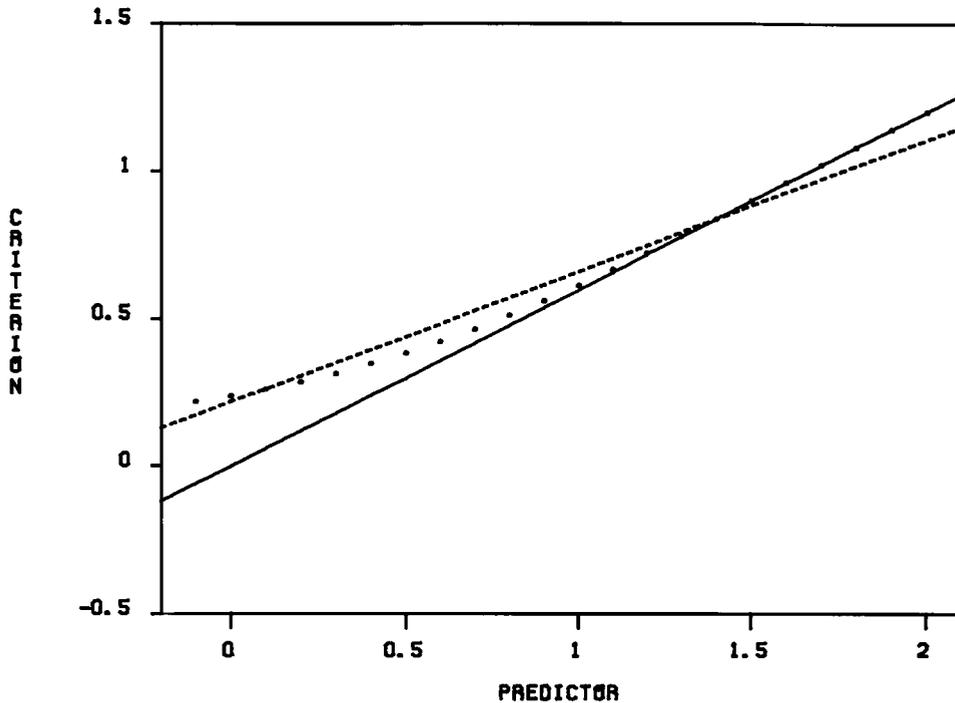
The dashed straight line in Figure 2 is the regression that would ordinarily be estimated in practice. However, the appropriate regression is represented by the solid line. Use of the dashed line in place of the solid line yields predictions that are too high for persons with low scores on X , i.e., it would lead to overprediction for an unselected low scoring group.

The availability of applicant data as well as data for matriculated students in the law school Validity Study Service files makes it possible to investigate the effects of selection on the regressions using one of several models (e.g., Heckman, 1979; Muthén & Jöreskog, 1983). Heckman has presented a two-stage procedure to estimate the regression coefficients in the unselected population. His approach to the problem starts with a probit analysis of the applicant data with the dichotomous criterion of accepted vs. not accepted. The probit analysis yields an estimate of $f(\lambda)$ and this estimate is used as an additional predictor. The regression coefficients of interest are then estimated using an ordinary least squares (OLS) regression of the criterion on the predictors and $f(\lambda)$.

In this study the Heckman procedure was used in one of the four schools with a sample of more than 100 black students. The procedure was applied separately to black and white students. The estimated regression coefficients based on OLS and the Heckman procedure are listed in Table 2.

Each estimation procedure yields weights for the predictors that are reasonably similar for the two groups of students. That is, the OLS regressions are nearly parallel to each other. So too are the planes based on the Heckman estimates. However, the weights for the predictors are considerably larger for the Heckman estimates than they are for the OLS estimates. Furthermore, the difference between the intercepts for white and black students is much smaller for the Heckman equations than for the OLS equations. In other words, the apparent overprediction for black students by the equation based on the sample of white students that is present in the OLS analysis almost disappears if the Heckman results are used.

Figure 2
 Linear Regressions in Unselected and Selectable Populations
 and Conditional Means in Selectable Populations



The results for this school are quite similar to what would be expected based on the hypothetical results presented in Figure 2. This can be seen most clearly by comparing the differences in the predicted FYA for OLS regressions for white and black students with the corresponding differences for the Heckman equations.

The differences in predictions for the OLS and Heckman equations are listed in Table 3 for values of the predictors equal to the means for black students and the means plus and minus one standard deviation for black students. The resulting nine combinations of predictor values span the region of predictor scores of the majority of the black students. As can be seen, the OLS equation overpredicts for black students between 3.11 and 4.31 in this region of predictor scores, whereas the amount of overprediction according to the Heckman results is between .08 and 1.01.

Discussion

It is clear that statistical artifacts can distort regression equations in ways that give the appearance of overprediction for a minority group. Selection on variables not included in the regression model can result in biased estimates of the regression parameters. It was argued that the bias may reasonably be expected to be in the direction consistent with the common finding of overprediction. Adjustment procedures such as Heckman's are intended to remove the bias due to selection effects. The results presented above from one school were consistent with expectations in that the Heckman procedure substantially reduced the amount of overprediction (from about .3 to .4 of a standard deviation for white students to between 0 and .1).

Although the Heckman procedure is potentially useful in studies of predictive bias, it is not without difficulties. The procedure requires additional as-

Table 2
Ordinary Least Squares and Heckman
Estimates of Regression Coefficients

Procedure and Sample	Regression Weights		
	Intercept	LSAT	UGPA
OLS			
White	16.2	.392	.284
Black	13.6	.375	.260
Heckman			
White	.6	.482	.434
Black	-.4	.470	.454

Table 3
Differences Between Predictions Based
on the Regression Equation for White Students
and the Regression Equation for Black Students at
Selected Values of the Predictors

Estimation Procedure	UGPA Value*	LSAT Value*		
		M-S	M	M+S
OLS	M+S	3.67	3.39	3.11
	M	3.99	3.71	3.44
	M-S	4.31	4.03	3.76
Heckman	M+S	1.01	.81	.62
	M	.74	.54	.35
	M-S	.48	.28	.08

*Values of predictors are (1) the mean for black students plus one standard deviation (M+S), (2) the mean for black students (M), and (3) the mean for black students minus one standard deviation (M-S).

sumptions and the obtained estimates have much larger sampling errors than their OLS counterparts (Dunbar, 1982). Hence, there is a tradeoff between consistency and sampling error. Nonetheless, the above results lend credence to the conclusion that the common finding of overprediction may be largely attributable to statistical artifacts.

References

- Dunbar, S. B. (1982). *Corrections for sample selection bias*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Evans, F. R. (1977). Applications and admissions to ABA accredited law schools: An analysis of national data for the class entering in the fall of 1976. In Law School Admission Council, *Annual Council Report*. Washington DC: Law School Admission Council.
- Heckman, J. J. (1979). Sample selection bias as specification error. *Econometrics*, 47, 153-161.
- Lawley, D. (1943). A note on Karl Pearson's selection formulae. *Royal Society of Edinburgh, Proceedings, Section A*, 62, 28-30.
- Linn, R. L. (1975). Test bias and the prediction of grades in law school. *Journal of Legal Education*, 27, 293-323.
- Linn, R. L. (1982). Ability testing: Individual differences and differential prediction. In A. K. Wigdor & W. R. Garner (Eds.), *Ability testing: Uses, consequences and controversies Part II*. Washington DC: National Academy Press.
- Linn, R. L. (1983). Predictive bias as an artifact of selection procedures. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement: A Festschrift for Frederic M. Lord* (pp. 27-40). Hillsdale NJ: Lawrence Erlbaum.
- Linn, R. L., & Werts, C. E. (1971). Considerations in studies of test bias. *Journal of Educational Measurement*, 8, 1-4.

- Muthén, B., & Jöreskog, K. G. (1983). Selectivity problems in quasi-experimental studies. *Evaluation Review*, 7, 139–174.
- Powers, D. E. (1977). Comparing predictions of law school performance for black, Chicano, and white law students. In Law School Admission Council, *Annual Council Report*. Washington DC: Law School Admission Council.
- Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, 36, 1128–1137.
- Schrader, W. B., & Pitcher, B. (1973). Predicting law school grades for black American law students. In Law School Admission Council, *Annual Council Report*. Washington DC: Law School Admission Council.
- Schrader, W. B., & Pitcher, B. (1974). Prediction of law school grades for Mexican American and black American students. In Law School Admission Council, *Annual Council Report*. Washington DC: Law School Admission Council.

Acknowledgments

A previous version of this paper was presented at an Invited Symposium: State-of-the-Art Series—Research on Bias in Testing, Annual Meeting of the American Psychological Association, Anaheim CA, August 27, 1983.

Author's Address

Send requests for reprints or further information to Robert L. Linn, College of Education, 210 Education Building, University of Illinois, 1310 South Sixth Street, Champaign IL 61820, U.S.A.