Affirmative Action, Duality of Error, and the Consequences of Mispredicting the Academic Performance of African American College Applicants

Jeryl L. Mumpower
Radhika Nath
Thomas R. Stewart

Abstract

The implications of different potential affirmative action policies depend on three factors: selection rate from the applicant pool, base rate of qualified applicants, and accuracy of performance predictions. A series of analyses was conducted under various assumptions concerning affirmative action plans, causes of racial differences in average college admissions test scores, and racial differences in accuracy of performance predictions. Evidence suggesting a lower level of predictive accuracy for African Americans implies that, under a program of affirmative action, both proportionately more false positives (matriculated students who do not succeed) and proportionately more false negatives (rejected applicants who could have succeeded) will be found among African American applicants. Unless equivalent levels of predictive accuracy are achieved for both groups, no admission policy can be fair simultaneously to majority group applicants and African American applicants. © *2002 by the Association for Public Policy Analysis and Management.*

INTRODUCTION

Vast amounts have been written over the past few decades on affirmative action (for summaries of recent research and extensive reviews of past research, see Bowen and Bok, 1998; Jencks and Phillips, 1998). It may well seem that little more is left to be said. Hammond's (1996) recent work integrating research on judgment and decision making with public policy, however, builds on a 60-year-old idea (Taylor and Russell, 1939) that can help to illuminate some neglected but important aspects of affirmative action policy.

The central concept in Hammond's argument is duality of error. This refers to the two possible types of mistake—false positives and false negatives—that may be made whenever policy decisions involve dichotomous choices, such as whether to admit or reject from college, accept or reject claims for welfare benefits, and so on. According to Hammond, duality of error inevitably occurs whenever decisions must be made in the face of irreducible uncertainty, or uncertainty that cannot be reduced by any activity at the moment action is required. (For similar arguments, see Connolly, 1987; Einhorn and Hogarth, 1981; Pauker and Kassirer, 1980.)

Hammond (1996) argues that any policy problem that involves uncertainty, and thus possesses the potential for dual error, will lead to unavoidable injustice. The college admission process will unavoidably reject qualified candidates. On the other hand, it will also accept applicants who fail to do the work. Calabresi and Bobbitt

Journal of Policy Analysis and Management, Vol. 21, No. 1, 63–77 (2002)

Published by John Wiley & Sons, Inc.

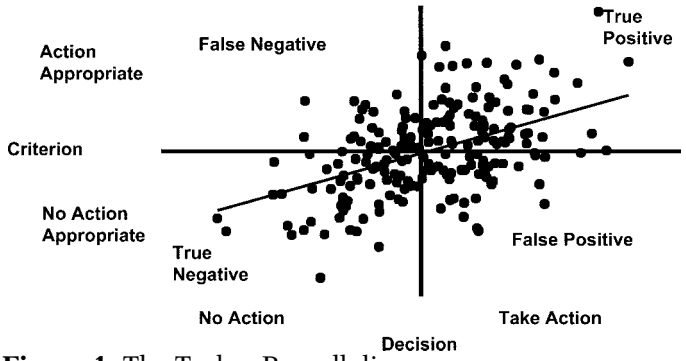


Figure 1. The Taylor–Russell diagram.

(1978) used the phrase "tragic choices" to describe similar policy problems in which perfect outcomes are impossible—seemingly, decision makers are damned if they do and damned if they don't. This paper demonstrates how duality of error leads to unavoidable injustice no matter what affirmative action policy, or absence thereof, is adopted.

The analysis discussed here relies on the Taylor–Russell framework (Figure 1), which illustrates the tension between the two types of error inherent whenever decisions based on imperfect information must be made about whether to take action. Each point in the scatterplot represents an individual case. The horizontal axis represents a judgment or score that is the basis for decision. If the score reaches a certain cutoff value, then positive action is taken (e.g., if the threshold is exceeded, admit the applicant to college). If the score does not reach the cutoff, no such action is taken (e.g., do not admit). The vertical axis represents the criterion. For individuals who fall above the criterion (represented by the horizontal line), the action is appropriate or correct. For individuals who fall below the criterion, the action is not appropriate or correct.

Decisions about applicants can be divided into four exhaustive and mutually exclusive categories. In addition to the two possible bad outcomes—false positives (Type I error: admit unqualified applicants) and false negatives (Type II error: reject qualified applicants)—there are two possible good outcomes—true positives (admit qualified applicants) and true negatives (reject unqualified applicants). The concepts of Type I and Type II errors date back to the first half of the 20th century (Neyman and Pearson, 1933, 1936) and are now a common part of statistical parlance. The incremental contribution of the Taylor–Russell framework is that it makes explicit the necessary interrelationships in dichotomous choice situations between Type I errors, Type II errors, and correct classifications.

The Taylor–Russell framework is also similar in some respects to the standard decision analysis table format for representing decision problems. With the four possible outcomes placed in a decision analysis table, and a corresponding payoff matrix, an expected value can be derived for any particular decision strategy (i.e., establishing a cutoff score for admission decisions). The Taylor–Russell framework enriches and extends the decision analysis approach, again because it makes clear necessary interrelations. The numbers in each cell are not independent, but rather depend simultaneously on the cutoff value, the criterion value, and the level of uncertainty. It is impossible for the number in any one cell to change without

necessitating a change in the value of at least one other cell. In most circumstances, reducing the number of false positives will increase the number of false negatives, and vice versa. Decreasing the number of mistakes of one type will typically lead to roughly proportionate increases in the number of mistakes of the other type.

Finally, the Taylor–Russell framework is also related to signal detection theory (Hammond, Lewis, and Hastie, 1992; Swets, 1973; Swets et al., 1991). For a given criterion, the data represented in the Taylor–Russell diagram could be analyzed using signal detection theory, and standard representations such as the receiver operating characteristic curve and the comparison of "noise" and "signal" distributions. Discussion in the present paper relies on the Taylor–Russell framework because, first, it deals explicitly with the situation where the prediction and the thing being predicted are both continuous and, second, it makes clear the relation between predictive accuracy (defined below) and error.

ERRORS IN PREDICTING FUTURE ACADEMIC PERFORMANCE

Decisions to admit students to undergraduate institutions typically consider, among other factors, high school grades, test scores from the Scholastic Aptitude Test (SAT) or ACT (American College Tests), recommendations, essays, and records of extracurricular activity. None is a perfect predictor of future academic performance, either singly or in combination.

Extensive empirical evidence indicates that predictions based on standardized tests typically over-predict the actual performance of African American students. Vars and Bowen (1998, p. 458) summarize their findings as follows:

> [W]hile SAT scores are related to the college grades of both black and white students, the relationship is weaker for blacks than for whites.... [A]t every level of SAT score, blacks earn lower grades than their white counterparts, and this remains true after controlling (at least crudely) for other variables, including high school grades and socio-economic status.... [T]he performance gap is greatest for the black students with the highest SATs.

Klitgaard (1985) explores in detail the implications of this over-prediction for highly selective institutions.

The magnitude of the over-prediction effect is substantial Figure (2); figure 2 summarizes data reported by Vars and Bowen. At six universities, African American students with SAT scores below 1000 have cumulative grade point averages (GPAs) approximately 0.27 grade points lower than majority-group students with the same SAT scores. For students with SAT scores exceeding 1300, the discrepancy between majority group and African American students widens to approximately 0.53 grade points. Vars and Bowen's finding that black students do not perform as well as white students with the same SAT scores is consistent with previous research (e.g., Breland, 1979), as is their finding that the gap widens at higher levels of SAT scores (Nettles, Thoeny, and Gosman, 1986).

For purposes of the following analyses, the pattern of over-prediction is important. If the gap between African American and majority group GPAs were constant across all levels of SAT scores, regression lines for the two groups would differ in terms of intercept, but would be parallel. (Parallel slopes might reflect a systematic bias in test scores or indicate that certain characteristics of the academic or social environment uniformly depress the performance of all African American students.) Assuming that the standard deviation of the two groups' SAT scores and GPA are roughly equivalent, parallel slopes would imply identical correlation coefficients between test scores and GPA for the two groups. The

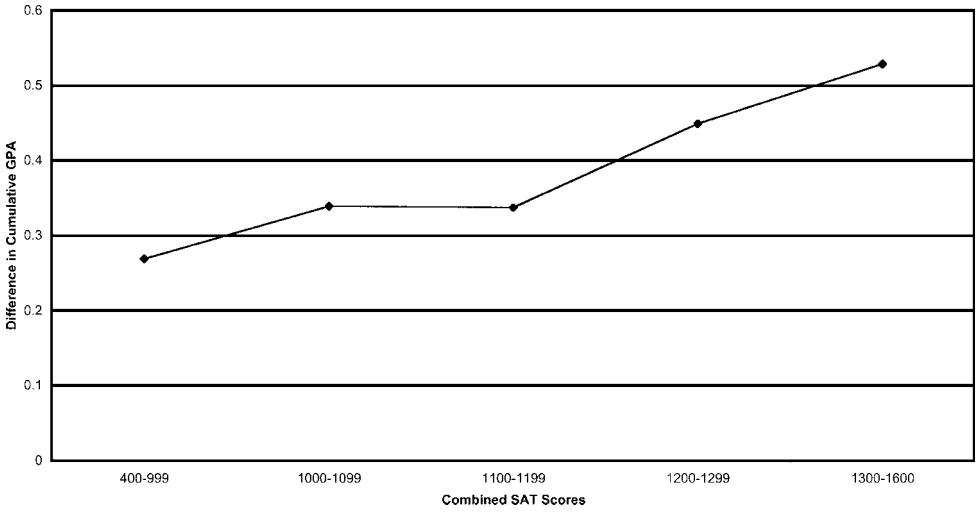


Figure 2. Mean differences in GPA between white and black students with same SAT scores (n = 14,457; based on data reported in Vars and Brown, 1998).

Vars and Bowen data, however, indicate differences in slope as well as intercept, which implies different correlation coefficients between predictor and dependent variables for the two groups, again assuming similar standard deviation of SAT scores and GPAs.

The validity and fairness of various potential predictors of academic success for white and black students has been a subject of intense study for decades (Breland, 1979; Hunter and Schmidt, 1976; Schmidt, Berner, and Hunter, 1973; Schmidt and Hunter, 1974, 1982; Vars and Bowen, 1998). Although the results have been somewhat mixed, it now appears fairly clear that decision-makers' ability to predict African American students' academic performance is less good than their ability to predict majority-group students' academic performance. Indeed, their ability to predict the academic performance of African American students is even poorer than that suggested by the predictive relationship between SAT scores and GPA, because their ability to predict who will graduate is poorer for African American students than for majority-group students. At every level of SAT score, graduation rates at highly selective institutions are significantly lower for African American students (Vars and Bowen, 1998).

The reasons for the phenomenon are less clear than its reality, as Vars and Bowen note. Measurement error represents one potential explanation of differential slopes across racial groups, as Phelps (1972) proposed in the context of labor markets. It would not be unreasonable to expect greater amounts of measurement error in SAT scores for African Americans, or any other non-majority subgroup, and such error normally would lead to lesser levels of predictive power. Vars and Bowen (1998) tested this hypothesis by comparing the test-retest reliability for majority group versus African American test-takers who took the SAT more than once. They found no evidence of greater measurement error for African American test-takers.

Given the attention the College Board pays to psychometric issues in test development and the fact that the large number of items contained by standardized

tests make high levels of reliability nearly inevitable (Nunnally, 1978), this finding is not surprising. Furthermore, measurement error alone cannot produce the effect shown in Figure 2. Measurement error would produce over-prediction for African American students relative to majority-group students at high SAT scores and under-prediction at low SAT scores. Figure 2 suggests either a non-linear relation between SAT scores and GPA or a combination of slope and intercept differences (i.e., a combination of measurement problems and treatment effects).

Perhaps standardized tests fail to measure some important predictive variable for African American students—a variable that is differentially relevant for high- versus low-scoring SAT takers. Alternatively, perhaps some unknown social or environmental variable adversely and disproportionately affects the performance of the very best African American students.

AFFIRMATIVE ACTION POLICY IN UNDERGRADUATE COLLEGE ADMISSIONS

Differential levels of predictive accuracy across racial groups have profound implications for college admission decisions. In the Taylor–Russell framework, three key factors influence any such analysis: the selection rate, or percentage admitted; the base rate, or percentage of those applying who could do the work if admitted; and predictive accuracy, or degree of correspondence between predictions of performance and actual performance. Ultimately, policies must be evaluated in terms of a fourth critical factor: the costs associated with false positives and false negatives, as well as the benefits associated with true positives and true negatives.

Unfortunately, few discussions on this or similar public policies consider all four factors. This means that implicit assumptions about omitted factors must be made. These buried assumptions may give rise to debates and disputes that will be difficult to resolve, unless they are brought to the surface and explicated.

Selection Rate

Colleges vary widely in the percentage of applicants they admit, ranging from highly selective institutions that admit fewer than 25 percent of their applicants, to community colleges that admit virtually every applicant. The present analysis concerns a hypothetical college that admits approximately 30 percent of its applicants—one that would be characterized as "highly selective" in most taxonomies. Different numerical results would be obtained for colleges with much more or much less selective admission policies. Analyses based on the Taylor–Russell framework are equally applicable to these other types of schools, but the results of the present analysis only generalize to schools that are highly selective.

Base Rate

The base rate corresponds to the percentage of applicants who are qualified to attend: What percentage of applicants would succeed if they were admitted In the present analysis, success is defined as graduation. Clearly, there are other possible definitions. Moreover, graduating from a highly selective college is not the same as graduating from a selective one, and neither of these is the same as graduating from a college with an open admission policy. Defining success in terms of graduation has two attractive virtues, however. First, graduation from college is a clear and commonly accepted measure of academic success. Second, it parallels the dichotomous nature of matriculation.

Because this analysis is specific to a particular college, any qualitative differences between what it means to graduate from different colleges or different types of colleges can be ignored. Neither does this paper address the fact that many students rejected from highly selective colleges go on to do well at less selective ones. Although rejection from the most selective colleges may have relatively little effect on an individual's overall welfare, this does not imply that the social consequences of unbalanced racial, ethnic, and gender composition in elite institutions are negligible.

The difficulties of estimating base rate are further confounded because fewer than half of admitted students matriculate at the typical selective college. Often, this is because they accept enrollment at another college of similar or better quality. This would argue for the conclusion that the ability of those who are accepted but do not attend is, on average, somewhat greater than those who do matriculate. For the present analysis the authors make the conservative assumption that, on average, the level of ability for those who were admitted and did not matriculate is precisely the same as those who were admitted and did matriculate.

For purposes of analysis, half of the applicant pool to this hypothetical highly selective institution is assumed to be capable of graduating. By capability is meant that the applicant possesses the combination of intellectual skills and personal characteristics that will lead to graduation. Clearly, this is merely an estimate. Although base rate is a critical notion, the type of experimental study that would produce significantly more precise estimates of this parameter—admitting students randomly—is unlikely ever to be performed. The analysis will consider both the possibility that there are no differences between racial groups in base rate, as well as the possibility that observed differences between white and African American racial groups in standardized test score results reflect real differences in preparation or capability.

Predictive Accuracy

It is difficult to establish an appropriate estimate of predictive accuracy for at least two reasons. First, there is a question about which measure to use—test scores, high school GPA, or some composite. Second, available estimates are based on a highly truncated sample. Not only do applicants self-select when they apply to institutions, the reported correlations between potential predictors (e.g., test scores) and performance are necessarily based solely on data from admitted and matriculated students.

Two key points influenced assumptions about levels of predictive accuracy in the present analysis. First, it is possible to obtain multiple correlation coefficients of at least 0.5 when predicting scholastic performance at highly selective colleges on the basis of multiple, commonly used predictors, including test scores, high school GPA, race, gender, and field of study (Vars and Bowen, 1998). Second, as discussed, African American students have repeatedly been found to have lower levels of achievement than would be expected on the basis of test scores, GPA, or other potential predictive measures (Vars and Bowen, 1998).

The present analysis assumes that an extremely good composite measure is available for majority-group students. For the sake of simplicity, this composite predictor will be referred to hereafter as a "test." Specifically assumed is a correlation between the test and academic performance of 0.8, when corrected for attenuation. Such an assumption is consistent with obtaining a multiple correlation coefficient of 0.5 based solely on the subset of applicants who matriculated at highly selective colleges. In all likelihood, it probably approximates an upper limit on predictability.

Consistent with the finding that standard predictive measures typically over-predict African American students' performance and that the difference is more pronounced at higher levels of SAT scores, it is assumed here that the correlation between the predictor and criterion for African American students is not as good as for majority students. Specifically, it is assumed that for this group, the correlation between predictor and academic performance is 0.65—a level of predictive ability that generally would be regarded as still highly respectable in behavioral domains.

Assessing Costs and Benefits

Conceivably, dollar values could be associated with the four cells of the Taylor–Russell diagram. In some public policy problems, assigning dollar values is fairly straightforward, but it would not be easy in the present instance. Surely, a variety of negative consequences are associated with failing to admit qualified applicants to competitive institutions. Qualified persons who are denied admission may lose earning potential, but estimating the magnitude of such losses is difficult. Qualified students narrowly denied admission to highly competitive institutions are likely to attend somewhat less competitive ones. The necessary data for estimating the magnitude of the losses that accrue from this outcome, to the individual and to the society, are simply not readily available. Further compounding estimation difficulties, these types of losses will almost always involve a combination of both private and public components.

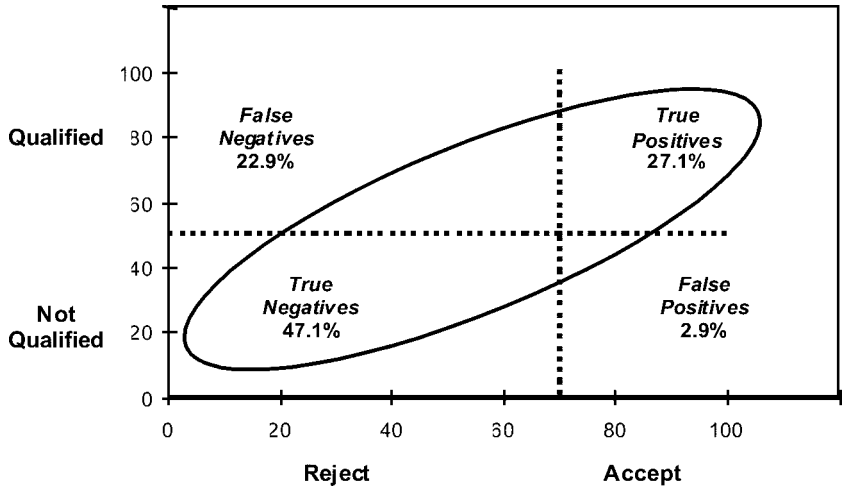
An alternative approach would be to use a decision analytic approach to assign utility rather than monetary value to outcomes. Although the authors are quite sympathetic to such approaches, the appropriate valuation of outcomes is a topic that engenders substantial debate, with little hope of resolution. For this reason, specific quantitative values have not been associated with any of the outcomes. This means the tradeoffs between the costs of errors and benefits of correct classifications will be discussed only in general terms.

Analysis of Undergraduate Admissions

For purposes of analysis, it is assumed that test scores and the criterion have a bivariate normal distribution. Results of the analysis for majority-group students are depicted graphically in Figure 3. Overall, 30 percent of majority-group applicants will be admitted. Of these, 27.1 percent are qualified applicants who are admitted; 2.9 percent are unqualified applicants who are admitted; 47.1 percent are unqualified applicants who are rejected; and 22.9 percent are qualified applicants who are rejected.

The probabilities shown in Figure 3 have been used to compute three relevant conditional probabilities for the majority group. These are shown at the bottom of column 1 of Table 1. The conditional probability that an admitted student is qualified is 0.90—an estimate that closely approximates the graduation rate for majority-group students at highly selective institutions. The probability of a qualified majority group student being admitted is 0.54, which means that only a little over half of all qualified applicants are admitted. The probability that an unqualified majority-group student is admitted is 0.06.

Columns 2 through 9 of Table 1 summarize a series of analyses regarding minority-group applicants, which can be contrasted to the base case analysis for majority group applicants presented in column 1. The logic of the analysis is a straightforward 2 x 2 $\times$ 2 design. The first factor concerns two different potential explanations for the fact that, on average, African American students' scholastic assessment test scores are



predictive accuracy = 0.8; base rate = 50 percent; acceptance rate = 30 percent

Figure 3. Taylor–Russell diagram for the admission of majority group applicants to highly selective colleges.

approximately one standard deviation below that of majority-group test-takers. This difference in test scores may reflect a genuine difference that results as a consequence of some combination of environmental or social causes, identified as the real difference condition. If so, about 16 percent of African American applicants are capable of doing the work required for graduation from this hypothetical highly selective institution. In contrast, the difference in test score may reflect systematic error. Cultural or other test biases may penalize African American test-takers by one standard deviation in their test scores. If so, then 50 percent of such test-takers are qualified, despite lower test scores.

The second factor depends on whether a program of affirmative action is in place. If there is no such program, then the only minority-group applicants who are admitted are those whose test scores exceed the same threshold for admission as that required for majority-group applicants. This means that a much lower percentage of minority-group applicants are accepted in comparison to majority-group applicants. Operationally, affirmative action programs may take a wide variety of forms. The present analysis assumes that affirmative action is implemented by adding a constant equal to the value of one standard deviation to the test scores of all minority-group applicants. As a consequence, minority applicants are accepted in exactly the same proportion as are majority group-applicants.

The third factor is the assumed quality of the admission test. In line with the previous discussion, the analysis first assumes that the predictive accuracy of the admission measure for minority-group applicants is 0.65. All analyses are then repeated, assuming that a test measure becomes available that offers the same high level of predictive accuracy (0.8) for minority-group applicants as for majority group applicants.

In column (2) of Table 1, results are given for the scenario that assumes a real difference between groups, no affirmative action, and use of an admissions measure with lower levels of predictive accuracy. Under this scenario, only 6.4

Table 1. Comparison of majority group and minority applicants to highly seletive colleges.

		Majority Group	Minority Group							
			Real Difference (16% assumed to be qualified)				Systematic Error (50% assumed to be qualified)			
		50% assumed qualified	No Affirmative Action		Affirmative Action		No Affirmative Action		Affirmative Action	
		predictive accuracy = 0.80	predictive accuracy = .65	predictive accuracy = .80	predictive accuracy = .65	predictive accuracy = .80	predictive accuracy = .65	predictive accuracy = .80	predictive accuracy = .65	predictive accuracy = .80
		(1)	*(2)*	(3)	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Column ID										
	Admitted	30.0%	6.4%	6.4%	30%	30%	6.4%	6.4%	30%	30.0%
	true positive	27.1%	4.2%	5.2%	11.5%	13.5%	6.0%	6.3%	24.5%	27.1%
Summary	false positives	2.9%	2.2%	1.1%	18.5%	16.5%	0.4%	0.1%	5.5%	2.9%
statistics	true negative	47.1%	81.8%	82.9%	65.5%	67.5%	49.7%	49.9%	44.5%	47.1%
	false negatives	22.9%	11.8%	10.8%	4.5%	2.5%	44.0%	43.7%	25.5%	22.9%
Conditional	p(qualified|admitted)	0.90	0.65	0.82	0.38	0.45	0.95	0.99	0.82	0.90
probabilities	p(admitted|qualified)	0.54	0.26	0.33	0.72	0.84	0.12	0.13	0.49	0.54
	p(admitted|not qualified)	0.06	0.03	0.01	0.22	0.20	0.01	0.00	0.11	0.06

percent of minority applicants are admitted, compared with 30 percent of majority-group applicants.

Not only are minority-group applicants admitted in proportionately lower numbers, under this scenario, qualified minority-group applicants are much less likely to be admitted than their majority-group counterparts. The probability that a qualified minority-group applicant is admitted is only 0.26, as compared with a probability of 0.54 for majority-group applicants.

Despite their lower admission rate, because the test measure has a lower level of predictive accuracy for minority-group applicants than for majority-group applicants, the proportion of false positives for minority-group applicants is greater than for majority-group applicants. Only 65 percent of admitted minority-group applicants are qualified, in comparison with 90 percent of admitted majority-group applicants.

The chances that an unqualified applicant will be admitted is small for both groups, but the chances are proportionately much greater that an unqualified majority-group student will be admitted than an unqualified African American applicant. The probability that an unqualified minority-group applicant will be admitted is only 0.03, in comparison with a 0.06 probability that an unqualified majority-group applicant will be admitted.

Under the assumptions of this scenario, it is less likely for African American students to be admitted than white students, generally. This holds true for both qualified and unqualified students. Among admitted students, however, it is proportionately more likely that minority-group students are unqualified and will fail to graduate.

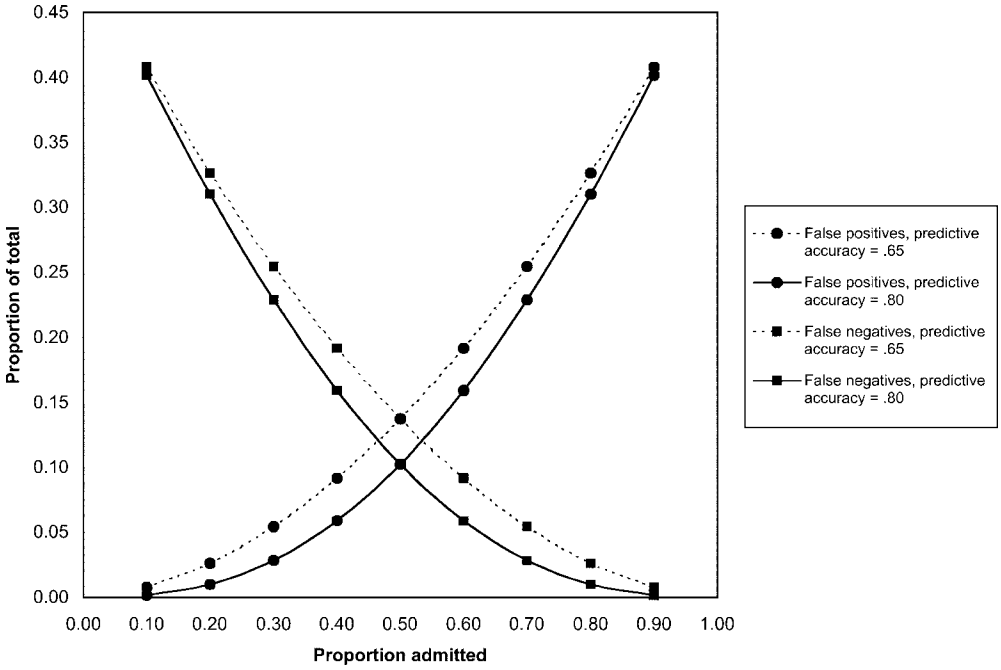
Columns 3 through 9 describe analyses based on the other seven combinations of assumptions. The various summary statistics and conditional probabilities provide different ways of looking at the tradeoff between false positives and false negatives. The pattern of results in the table is illustrated by the next to last row, showing the results for the conditional probability of being admitted, given that the applicant is qualified.

A major argument for affirmative action is that, without it, qualified minority applicants are less likely to be admitted than are qualified non-minority applicants. Comparing $p(\text{admitted}|\text{qualified})$ for column 1 with columns 2 and 3 and 6 and 7 shows that this is clearly the case. But columns 4 and 5 (which assume that only 16 percent of minority applicants are qualified) support the argument of some affirmative action opponents that the inequity may be reversed when affirmative action is implemented. However, when it is assumed that minority and non-minority applicants are qualified in equal percentages (columns 6–9), affirmative action alone does not compensate for the inequity unless the predictive accuracy of the admission test is equal to that for non-minorities (column 9). In other words, the success of affirmative action in compensating for admissions inequity depends on the predictive accuracy of the admissions test.

Assumptions of the Analysis

No doubt, none of the eight scenarios examined capture the true state of affairs accurately and completely. Collectively, however, they demonstrate the critical importance of paying greater attention to the significant, potentially pernicious effects of differences in decision-makers' ability to predict majority- versus minority-group students' academic performance.

Numeric results are highly dependent on specific assumptions concerning the percentages of applicants who are qualified, the percentage who are admitted, and predictive accuracy. The degree of sensitivity to assumptions is illustrated in Figure



Note: assuming 50 percent to be qualified.

Figure 4. Proportion of error as a function of predictive accuracy and proportion admitted.

4, which gives the predicted percentages of false positives and false negatives, as a function of the percentage admitted, assuming that 50 percent of applicants are qualified, and varying the assumed predictive accuracy from 0.65 to 0.80. As can be seen, the proportion of false positives and false negatives over these conditions range widely, with false negatives outnumbering false positives when greater selectivity is assumed (and vice versa when lesser levels of selectivity are assumed). Both types of errors are lower when predictive accuracy is greater, regardless of the proportion admitted. Many more such analyses of assumptions are possible than can be presented here.[1]

The Effects of Affirmative Action Programs on Non-minority Applicants

Historically, one of the most troublesome and divisive issues has been the view that if the number of admission slots is limited, at least some majority-group applicants who would otherwise have been admitted will be refused admission if minority-group admissions are increased as a result of affirmative action. Although the overall magnitude of the effect may be small, the implications of affirmative action for rejected majority-group applicants who would have been admitted otherwise (or believe this

[1] The authors are happy to provide access to the software used in these analyses to interested readers who wish to perform further analysis on their own. Please contact Thomas R. Stewart to make arrangements.

to be the case) has served as a persistent, highly charged emotional lightning rod for societal debates about affirmative action.

The magnitude of the effect of affirmative action programs on non-minority applicants depends on the size of majority- and minority-group populations, in addition to the base rate, admission rate, and the level of predictive accuracy. It is possible to address the question more specifically in the context of the above numeric example, but several additional assumptions must be made explicit. First, assume that 12.5 percent of applicants are African American, a percentage that is consistent with the current racial make-up of the United States. For purposes of simplicity, treat the remaining 87.5 percent as a single group, although there are in reality substantial and significant differences among admission rates for majority-group members, Asian Americans, Hispanic Americans, and Native Americans. In line with the analyses reported in Table 1, assume that with no affirmative action program in place only 6.4 percent of minority-group applicants are admitted, whereas 30 percent of majority-group applicants are admitted.

Unless the institution achieves affirmative action by increasing its capacity so that 30 percent of both majority- and minority-group students can be admitted, the number of majority-group applicants must be reduced. In the present numerical example, to maintain a constant size, the percentage of applicants admitted must equal approximately 27 percent for both groups. For minority applicants, this means raising the admission rate from 6.4 percent to 27 percent, for a greater than fourfold increase. For non-minority applicants, the admission rate must be lowered from 30 percent to 27 percent, approximately a 10 percent decline in the number of non-minority students admitted.

CONCLUSION

Debates about affirmative action have included much discussion about whether the factors used in college admissions decisions, particularly standardized tests, are systematically biased such that they underestimate the ability and potential of all minority-group applicants. The above analysis has demonstrates that another factor—differences in decision-makers' ability to predict accurately majority- and minority-group members' academic performance—creates serious, generally unrecognized complications for affirmative action policy. Because of differences between groups in decision-makers' ability to predict accurately, the ratio of false positives and false negatives will be different for majority- and minority-group applicants. Generally, this implies two unattractive options. Admitting minority-group applicants in the same proportion as majority-group applicants will create a higher proportion of false positives among minority students. But, to avoid the first problem, admitting a lower proportion of minority-group applicants than majority-group applicants will create a higher proportion of false negatives among minority students. Given the present level or predictive ability, decision-makers are likely to make more mistakes (of both types) in admission decisions for minority-group applicants than in admission decisions for majority-group applicants.

The current analyses should be regarded as primarily illustrative, although the authors believe the assumptions are plausible and defensible ones. For use in any specific policy context, analyses tailored to the specific institution in terms of acceptance rates, base rates, and perhaps even the degree of predictive accuracy will be required. The analyses must be tailored to the institution, and sensitivity analyses should be done to account for the uncertainty that will inevitably be associated with key parameters.

Further, it is important to recognize that the present analysis represents a one-generation model. Decisions about affirmative action policy and the consequences of such decisions for later generations are important topics not addressed here. If affirmative action works as its proponents hope, higher levels of minority admission today will lead tomorrow to higher rates of success among minority students. Conversely, critics argue that higher levels of admission today may backfire if they lead to higher rates of failure. Such positive or negative feedback across generations is critically important, but is not addressed by the current analysis.

The above analysis of the affirmative action policy leads to three general conclusions, with implications beyond this specific analysis.

First, without taking steps to improve predictive capability, changes in policies will primarily involve simply trading off one type of error for the other. Uncertainty creates a tradeoff between false positives and false negatives. Reducing one type of error leads inevitably to an increase in the other. The only way to reduce one kind of error without creating more of the other kind is to reduce the uncertainty. Methods for doing this include obtaining better and more reliable information for decision-making and making better use of the available information. In the interest of fairness, it is imperative to develop and use the best possible predictors of future performance. In short, until better tools have been developed for predicting the performance of minority-group applicants these tradeoffs among the two types of errors will plague decisionmakers and policymakers. Better prediction does not obviate the need for taking action to improve the capabilities of minority-group students. The graduation rate of minority-group students can be improved both by increasing the base rate (i.e., creating a larger pool of more capable students) and by improving predictive accuracy (i.e., developing tools that better predict whether or not applicants will succeed in school).

Second, the appropriate tradeoffs between false positives and false negatives lie at the crux of policy debates in many different substantive areas. Since the uncertainty that necessitates trading one kind of error for another can never be completely eliminated, choosing a course of action means assigning, implicitly or explicitly, relative values to the two kinds of error. In the present context, the difficulties are compounded because there will inevitably be different rates of the two kinds of errors for minority-group than for majority-group applicants.

Third, unjust or unfair treatment of specific persons or groups is inevitable for any policy that is applied in a context of irreducible uncertainty. Any course of action creates both "winners" and "losers." As much as people would like to eliminate all fairness and injustice, it is unavoidable. This injustice results not from the selfishness or cruelty or incompetence of policymakers or their policies, but simply from uncertainty. The effects are compounded when different levels of uncertainty are associated with different groups in society, resulting in inequitable distribution of winners and losers across groups.

In conclusion, debates about the design of affirmative action policies, their success, and their evaluation most often focus on the unfairness inherent either in rejecting qualified applicants or accepting unqualified ones. The linkage between these two types of errors is seldom if ever explicitly recognized. This analysis has shown that under normal circumstances taking steps to reduce one type of error will lead to increases in the other. No admission policy can be fair simultaneously to majority-group applicants and African American applicants, unless and until equivalent levels of predictive accuracy are achieved for both groups.

The authors wish to thank two anonymous reviewers of an earlier version of this paper for their helpful comments and suggestions.

JERYL L. MUMPOWER is Professor at the Center for Policy Research and Department of Public Administration and Policy, Nelson A. Rockefeller College of Public Affairs and Policy, State University of New York, Albany.

RADHIKA NATH is a Doctoral Student at the Center for Policy Research and Department of Public Administration and Policy, Nelson A. Rockefeller College of Public Affairs and Policy, State University of New York, Albany.

THOMAS R. STEWART is Research Professor at the Center for Policy Research and Department of Public Administration and Policy, Nelson A. Rockefeller College of Public Affairs and Policy, State University of New York, Albany.

REFERENCES

Bowen, W.G., & Bok, D.C. (1998). The shape of the river: Long-term consequences of considering race in college and university admissions. Princeton, NJ: Princeton University Press.

Breland, H. (1979). Population validity and college entrance measures (Research Monograph 8). New York.

Calabresi, G., & Bobbitt, P. (1978). Tragic choices. New York: W. W. Norton.

Connolly, T. (1987). Decision theory, reasonable doubt, and the utility of erroneous acquittals. Law and Human Behavior 11, 101–112.

Einhorn, H.J., & Hogarth, R.M. (1981). Behavioral decision theory: processes of judgment and choice. Annual Review of Psychology 32, 53–88.

Hammond, K.R. (1996). Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice. New York: Oxford University Press.

Hammond, K.R., Lewis, H.O., & Hastie, R. (1992). Making better use of scientific knowledge: Separating truth from justice. Psychological Science 3, 2, 80–87.

Hunter, J.E., & Schmidt, F.L. (1976). Critical analysis of the statistical and ethical implications of various definitions of test bias. Psychological Bulletin 83, 6, 1053–1071.

Jencks, C., & Phillips, M. (Eds.). (1998). The black-white test score gap. Washington, DC: Brookings Institution Press.

Klitgaard, R. (1985). Choosing elites. New York: Basic Books.

Nettles, M.T., Thoeny, R., & Gosman, E.J. (1986). Comparative and predictive analyses of black and white students' college achievement and experiences. Journal of Higher Education 57, 289–318.

Neyman, J., & Pearson, E.S. (1933). On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society, A 231, 289–337.

Neyman, J., & Pearson, E.S. (1936). Sufficient statistics and uniformly most powerful test of statistical hypotheses. Statistical Research Memoirs 1936, 1, 113–137.

Nunnally, J.C. (1978). Psychometric theory, 2nd ed. New York: McGraw-Hill.

Pauker, S.G., & Kassirer, J.P. (1980). The threshold approach to clinical decision making. New England Journal of Medicine 302, 1109–1117.

Phelps, E.S. (1972). The statistical theory of racism and sexism. American Economic Review 62, 659–661.

Schmidt, F.L., Berner, J.G., & Hunter, J.E. (1973). Racial differences in validity of employment tests: Reality or illusion. Journal of Applied Psychology 58, 1, 5–9.

Schmidt, F.L., & Hunter, J.E. (1974). Racial and ethnic bias in psychological tests: Divergent

implications of two definitions of test bias. American Psychologist 29, 1, 1–8.

Schmidt, F.L., & Hunter, J.E. (1982). Two pitfalls in assessing fairness of selection tests using the regression model. Personnel Psychology 35, 3, 601–607.

Swets, J.-A. (1973). The relative operating characteristic in psychology. Science 182, 4116, 990–1000.

Swets, J.A., Getty, D.J., Pickett, R.M., D'Orsi, C.J., Seltzer, S.E., & McNeil, B. J. (1991). Enhancing and evaluating diagnostic accuracy. Medical Decision Making 11, 1, 9–18.

Taylor, H.C., & Russell, J.T. (1939). The relationship of validity coefficients to the practical applications of tests in selection. Journal of Applied Psychology 23, 565–578.

Vars, F.E., & Bowen, W.G. (1998). Scholastic aptitude test scores, race, and academic performance in selective colleges and universities. In C. Jencks & M. Phillips (Eds.), The black-white test score gap (pp. 457–479). Washington, DC: Brookings Institution Press.