The Effect of Law School Racial Preferences on Minority Bar Performance

Project Description

I. Racial Preferences in Higher Education

As America moved from policies of racial exclusion to policies of inclusion in the 1960s, many hoped that declines in discrimination would produce greater integration and a convergence in social and economic outcomes between blacks and whites. Though this occurred in some settings, many observers in higher education concluded that progress there would be slow indeed if racial barriers to entry were simply dropped without any affirmative efforts to improve minority access and participation. One obvious strategy involved outreach (Toepfer 1966). But a fundamental obstacle for many schools was the underlying, pervasive gap between blacks and whites on standardized tests and other measures of academic achievement. On most such exams, such as the SAT and the LSAT, black scores lagged behind whites (usually by more than a full standard deviation). Somewhat smaller but still large gaps existed in conventional GPA measures as well. This meant that at elite colleges and many graduate programs, admissions based largely on hard credentials would produce only tiny numbers of black enrollees. Many schools consequently adopted policies in the late 1960s and early 1970s that essentially used different quantitative admissions standards for different groups (especially blacks) so that the racial/ethnic composition of the students admitted would more closely reflect the composition of the underlying applicant pool.

These large racial preferences were initially viewed as a temporary expedient (*Bakke*, 1978). A key reason for the preferences was to convey to minorities that elite educations really were open in a way they had not been before and, indeed, the minority applicant pool grew steadily. Most educators assumed that as other civil rights reforms took hold (especially reform of K-12 education systems), the black-white test score gap would rapidly decline, and that within a generation, racial preferences would be unnecessary to assure adequate representation. In fact, the black-white test score gap declined significantly in the 1960s and 1970s, but progress largely stopped around 1980, with the gap still on the order of a standard deviation (Jencks and Phillips, 1998). Recent research has shown compellingly many of the environmental influences that make up the gap (Fryer & Levitt, 2004). Among law school applicants, the gap has continued to narrow, but remains at more than a full standard deviation (Sander 2004). The size of the gap has prompted elite colleges and graduate schools to maintain admissions techniques that rely on some sort of race-norming.

The early 1990s saw the renewal of a national debate on whether large preferences should continue to be standard practice in much of higher education. On the legal and policy front, the debate gathered steam with *Hopwood* (an appellate decision that prohibited race-based admissions preferences in the Fifth Circuit), generated national attention with the passage of the California Civil Rights Initiative (Prop 209), and climaxed with the *Gratz* and *Grutter* decisions of 2003, which left education administrators with considerable discretion in their use of preferences to promote diversity on campuses. The continuing spread of statewide bans (to Washington State and Florida in 2000, and Michigan in 2006) is indicative of the continuing importance of the issue in public policy. As we shall see, the policy debate is increasingly entwined with a vigorous debate among social scientists about the impact affirmative action has on its intended beneficiaries.

II. The Mismatch Debate

Racial preferences in higher education have been based on two broad justifications: the benefits of increased diversity for everyone at universities, and the benefits of greater access to higher education (especially at its most elite levels) for racial minorities who receive preferences. For many years, the "diversity" rationale received more attention (and became the Supreme Court's central premise for permitting racial preferences). The benefits of preferences to minorities admitted under special programs tended to be taken for granted. Many educators simply equated preferences with access, and assumed preferences were beneficial because the implicit alternative, exclusion, was obviously bad. Others have noted that for many (if not most) minorities, preferences did not determine whether they could go to college or professional school at all, but rather where in the "eliteness" hierarchy they would land. Landing higher up in this hierarchy is generally assumed to be a good thing, for several reasons. A more elite degree often carries more weight in the job market (or in subsequent stages of education). At more elite institutions, students make more powerful connections (Wilkins, 2005). Elite schools tend to have more resources per student and more talented professors (at least in some dimensions), and this should help students. Last but not least, the actual immersion in an environment of more talented students is arguably a stimulus to higher achievement – an advantage, not a disadvantage (Hoxby 2000, Stinebricker & Stinebricker 2000, Sacerdote 2001, Zimmerman 2003, Winston and Zimmerman 2003).

Opposing this reasoning is the theory of "educational mismatch", articulated by early observers such as Thomas Sowell (1974) and Clyde Summers (1970), and discussed by many later critics of preference policies (e.g., Thernstrom 1997). The mismatch theory really involves two different types of argument. The first is that optimal learning will not occur if a student's own preparation and cognitive skills are substantially less developed than those of most of the student's peers, especially in competitive, fast-paced settings. Teachers may pitch their instruction at a level too difficult for the mismatched student to fully absorb; difficulty keeping up becomes more of a concern as the semester wears on, and a student will ultimately learn less and perform badly. The effect is similar to college freshmen trying to skip first-year physics to go straight into advanced classes: some students may learn faster, but many may crash and burn. (Note that this effect does not turn on race per se; any student facing a large credentials gap with his classmates is vulnerable.)

The second part of the traditional mismatch argument focuses on the group dynamics that may result from the creation of an environment in which entering credentials correlate strongly with race. If credentials even moderately predict school performance, then performance (e.g., grades) will correlate with race as well. This may have a variety of harmful effects: reinforcement or creation of negative stereotypes among whites, feelings of victimization among minorities, tendencies towards group self-segregation, and so on.

Although mismatch arguments occasionally surfaced in policy debates and more philosophical discussions of preferences (e.g., Steele, 1991), it was not until the past decade that they attracted significant (and increasing) attention from social scientists.

Several sociologists have framed the issue in terms of the "frog pond" analogy: is it better to be in a less competitive educational environment where one will shine (relatively), or to be in a more competitive environment where one may have an undistinguished record but absorb lots of human and social capital from one's fellow students? (Esplenshade et al, 2005) Several

sociologists have considered whether one's academic strength relative to one's peers might affect one's course of study and career plans in college. Elliott (1996) found that students who had credentials significantly lower than their classmates were far more likely than other students to abandon science majors. The intuition behind this finding is compelling: students with relatively lower credentials will find large, competitive, curve-graded science classes particularly intimidating, and will drift towards "softer" curricula. Cole and Barber (2003), approaching a different question with different data, obtained a very similar result: black students who faced a credentials gap with their classmates were much less likely to stick with an early interest in academia, when compared with black students admitted with smaller or no preferences. The intuition is, again, straightforward: students are less likely to want to become professors if they are not fairly successful as undergraduates relative to their college peers.

In recent years, economists have become quite interested in a closely related question: what sorts of higher educational environments best promote the acquisition of human capital? Questions are prompted in part by the very hierarchical structure of post-secondary education; nearly every field of higher education is marked by the sorting of students across institutions by ability, and there is a good deal of evidence that this sorting has increased in recent decades (Frank & Cook, 1993; Stake, 2006). One possible explanation of this sorting is that students learn best in college and graduate school when they are in classrooms with peers of similar ability or cognitive skill. Epple et al (2003) showed that if this "matching" hypothesis is true, then the sorting of students by cognitive skill across different schools would characterize any equilibrium that exists.

Economists Loury and Garman (1996) used the High School and Beyond database to explore how relative credentials affected performance. They found evidence consistent with the mismatch effect for both blacks and whites. Students with credentials well below their classmates' averages had lower grades, usually lower graduation rates, and usually lower earnings than students who were not mismatched, when credentials were held constant.

A much larger and far more discussed study appeared in 1998. In *The Shape of the River*, Bowen and Bok used specially developed data (known as the College and Beyond database) which described three large, longitudinal cohorts of students who had attended colleges ranging from moderately elite to very elite. Bowen and Bok found little evidence of a mismatch effect. Blacks and Hispanics who attended more elite colleges had higher graduation rates and higher earnings than did students with similar credentials who attended less elite schools. Bowen and Bok did find large grade disparities between blacks and whites (the average black GPA in their sample was at the 22nd percentile), but suggested that this was due to black underperformance as much as it was due to admissions preferences. Indeed, the authors argued that racial preferences at elite colleges in the 1980s and 1990s were relatively modest in many ways, and that race was generally not a determinative factor in admissions.

Critics like Thernstrom (1999) pointed to some of the problems with the Bowen & Bok analysis – problems which continue to bedevil much research on the effects of preferences. The most serious of these was selection bias. A typical Bowen and Bok regression looking at future earnings, for example, would classify each student as falling within one of three ranges of high school grades and one of four ranges of SAT scores. The regression would then, in effect, compare the long-term earnings of pairs of students with similar credentials (e.g., middle third of their high school class and SAT score between 1200 and 1299) but different colleges (e.g., Harvard vs. University of North Carolina), and find that the Harvard graduate earned more than

the UNC graduate. But in truth, the two students were probably not quite comparable. The Harvard student would be more likely to be near the top of the measured credential ranges (e.g., have grades near the top of the middle third and SAT scores close to 1299) than would the UNC student. And even if the students had identical numerical credentials, in all likelihood the Harvard student had other "unobserved" credentials that made him stronger than the UNC comparison student (e.g., a better high school, more AP courses, impressive application essays, etc.). Thus, the regression was biased towards finding better outcomes for Harvard students.

Dale and Kreuger (2004) demonstrated how important these selection effects could be in their reanalysis of the Bowen and Bok data. They found that it was possible, using the College and Beyond dataset, to match pairs of students who had been admitted to the same pair of institutions (e.g., again, Harvard and UNC) but had chosen to attend different schools (one to Harvard and one to UNC). This approach probably avoided selection bias, since the comparisons were between two students who were both strong enough to secure admission to the more elite institution. (Of course, this does not create a true experimental condition since student choices are not made randomly; it is possible that students who choose the more elite institution tend to have other attributes that make them stronger, or weaker, than their counterpart choosing the less elite school. But this at least vastly reduces selection bias.) Using several variations of this method, Dale and Krueger found either no effect or a negative effect on earnings for more elite students – just the opposite of the results found by Bowen and Bok.

Of course, the main premise of the "mismatch" hypothesis is that graduation rates and earnings are just indicators of a more fundamental measurement: how much students actually learn. Bowen and Bok relied on graduation rates as their primary "academic" outcome measure – and a good deal of the other mismatch literature discusses graduation rates. But graduation is a conspicuously imperfect measure. Each college has complete control over its own graduation rate, and many elite schools make great efforts to graduate all of their students, regardless of the student's actual academic achievements. Because college graduates no not take uniform exit exams, essentially none of the college literature measures learning, or attempts to measure mismatch effects upon learning.

III. Law School

Until recently, nearly all the empirical analyses concerning the effects of preferences (including those discussed above) had focused on undergraduate education. This changed with the release in 2001 of an unusual set of longitudinal data on law students, known as the Bar Passage Study (BPS) and created by the Law School Admissions Council (LSAC). LSAC developed the data not to study the effects of preferences *per se*, but to determine the extent to which bar exams (used throughout the U.S. to license lawyers) produced disparate outcomes by race. LSAC secured the participation of ninety percent of the nation's accredited law schools, and seventy percent of the students at those schools who matriculated in 1991. The BPS tracked these students for six years, creating a rich database on the credentials, experiences, grades, and bar results of 27,000 students.

The BPS revealed distressing patterns. Blacks were nearly two-and-one-half times as likely as whites to not graduate from law school, four times as likely to fail their first bar exam, and six times as likely to fail the bar after multiple attempts. In a widely-discussed 2004 article, Sander combined other data with the BPS to both present a detailed description of the operation of preferences and an argument that mismatch effects were large and relatively pervasive.

Setting aside a handful of historically black law schools, Sander showed that the gap in entering credentials between blacks and whites was very large at all law schools, regardless of a school's eliteness. The black-white gap in credentials at a typical law school generally appeared to be about two standard deviations or more (Sander 2004). This pattern was most easily explained by schools race-norming admissions standards. Moreover, incoming credentials proved to be unbiased predictors of performance in law school (i.e., black underperformance, contrary to Bowen and Bok's findings for college, was either nonexistent or minimal in law school[1]); the large racial gap in credentials at each school thus translated into a large gap in grades. Again excepting historically black law schools, the median black at nearly all other schools obtained grades at the 6th to 8th percentile of the white grade distribution. In other words, half of all black students ended up in the bottom tenth of their class.

The key question is whether the low grades blacks incur as a result of preferences are offset by the value of attending a more elite school. Which is more important, an elite school or good grades, in graduating from law school, passing the bar, and getting a good job? One could try to answer this question by measuring whether, when one controls for entering credentials, black law students are more or less likely to have good outcomes when they graduate from more elite schools – essentially the Bowen and Bok approach. The problem with this approach is the familiar one of selection bias; if one compares two blacks with LSATs of 155 and UGPAs of 3.4, it is likely that the student going to the more elite school is there for a reason – there are likely to be unobserved characteristics (e.g., attendance at a more elite college, following a tougher curriculum, demonstrating stronger writing ability) that make the more "elite" student academically stronger, independent of law school effects. The problem is greatly compounded by the fuzziness with which the BPS database measured eliteness. Rather than identifying individual schools, BPS classified them into one of six "clusters", organized partly, but not exclusively, by school eliteness (size, cost, and public/private status are also weighed). For example, it is quite plausible that small, private Vanderbilt Law School was classified in what is, on the whole, the most elite of the six clusters, while larger, public University of Michigan Law School may have been classified in the second or even the third most-elite cluster -- even though by nearly any school-by-school ranking, Michigan would rank far above Vanderbilt. For all these reasons, a simple regression on individual credentials would produce fuzzy and non-significant results.

To deal with these weaknesses in the data, Sander offered two more deductive arguments. First, he contended that whites and blacks with similar index scores had similar unobserved credentials; since most blacks received large admissions preferences and most whites did not, comparing the outcomes of whites and blacks after controlling for credentials provided a way of observing mismatch effects (and under this method, such effects appeared to be quite large). Second, Sander looked at the size of the grades/prestige tradeoff that resulted from admissions preferences, and then tried to measure how much the lower grades, and the higher prestige, each affected subsequent outcomes. He found that for graduation and bar passage outcomes, grades were far more powerful predictors than was school eliteness; for earnings outcomes, grades appeared to be more powerful at all but the most elite schools. Sander estimated that the lower grades resulting from preferences had a large enough net cost to blacks to account for roughly

---

[1] Research in the 1970s had found significant evidence of black underperformance in law school, but a comprehensive analysis by Anthony and Liu (2004) found "very slight" (about one-eighth of a standard deviation) black underperformance for more recent cohorts.

half of the black-white gap in bar passage rates; to put it differently, maximizing one's racial preference could roughly double a black law student's chances of eventually failing the bar. Of course, Sander's second argument was somewhat undermined by the fuzzy measure of school eliteness in the BPS; although grades seemed to be far more powerful than eliteness in predicting outcomes, this might be due to the far greater precision with which the database measured grades.

In a subsequent analysis, Sander (2005) used a different methodology to examine outcomes with the BPS data. A student questionnaire in the BPS asked matriculating law students about the process of applying to law school. About one-seventh of the black students who had been admitted to more than one law school reported they had turned down their "first choice" school to attend an apparently less elite school, usually for financial or geographic reasons. Examining black law students who had been admitted to a more elite school but had chosen a less elite one (known as "second choice" students) provided a way – similar in some respects to the Dale-Krueger method – of avoiding the problem of selection bias. With this method, there was a greatly reduced risk that differing outcomes between the two groups could be attributed to unobserved differences in incoming credentials. The technique also mooted complications of race, because one could compare blacks directly with other blacks. The results of this analysis lined up closely with Sander's earlier estimates. Blacks choosing to attend less elite schools had significantly higher grades and reduced their chance of failing the bar by nearly half – and the results held up with a wide variety of added controls.

These articles generated wide discussion and responses by over a dozen legal academics and economists. Several critics used Bowen & Bok-type regressions (without correcting for selection bias) to argue that Sander's results were misleading or at least ambiguous. Overall, three alternative views emerge from the literature and discussion of the law school mismatch research: (1) evidence of mismatch is substantial and persuasive (Sander, 2004 and 2005); (2) data is partly consistent with mismatch, but there may be other, undefined explanations (Rothstein and Yoon 2006: mismatch may exist for most black law students, but data problems preclude a definitive answer); (3) data is inconsistent with mismatch, but there is no good alternative explanation of the black-white disparities in outcomes (Chambers et al. 2005, emphasizing inconclusiveness of intra-black population regressions).

Of the various possible mismatch outcomes facing law students who receive preferences – on grades, graduation rates, bar passage, and future earnings – there is particular interest, for several reasons, in bar passage. The bar is taken by 94% of law graduates nationwide; passing it is almost everywhere a requirement for entry to the legal profession, and failure dramatically limits job opportunities. A number of states have toughened their bar exams in recent years (Merritt, 2001), and the national proportion of first-time bar takers passing the bar fell from 82% in 1994 to 75% in 2004. These trends have affected blacks disproportionately, especially in terms of ultimate bar passage; we estimate that among blacks starting law school in 2001, well over half did not become lawyers (compared to a fifth for whites), with the bar exam being the single greatest source of attrition.

Examining bar outcomes is interesting and important for another reason: a state bar exam represents a uniform effort to measure the learning of law graduates. It is thus a more direct approximation of what "peer effects" and "mismatch effects" are supposed to be about – the optimal learning environment for students. Bar examinations do not cover many things students study in law school, but they do cover many things all law students do study: knowledge of the

substantive law; skill in identifying relevant legal issues in the context of a legal problem; and skill in making a reasoned case for how legal principles resolve disputes. The consistently strong correlation between law school grades and bar scores (.7 to .8) is consistent with the claim that bar exams test knowledge and skills taught in law school. From a theoretical point of view, this relatively close link makes this setting unusually apt for examining whether mismatch effects actually affect learning.

IV. The Need for Better Data

Despite many disagreements over the extent of the mismatch effect, there is very wide agreement among legal scholars that better data is essential if real progress is to be made in understanding the reasons why blacks and Hispanics have so much difficulty with the bar (Rothstein & Yoon 2006; Barnes 2006; Sander 2005) The major shortcomings of the BPS are these:

1) The BPS collected information only on whether students passed the bar exam on each attempt; it provided no information about actual bar scores. Bar data is thus dichotomous, which makes inferences about performance quite crude (especially if, as in the BPS, seven-eighths of all bar-takers pass on their first attempt). Actual bar scores allow for standard OLS techniques and dramatically better measurement of effects.

This is particularly true in examining mismatch effects at the "top of the distribution." Some scholars have argued that mismatch effects are non-existent or "reversed" for beneficiaries of preferences at the most elite schools (Rothstein & Yoon, 2006; Chambers, 2005). But such a conclusion could follow fallaciously from dichotomous data: if the racial gap in bar scores is identical at elite schools but the absolute scores are higher, the difference in bar passage rates would at some point decline into statistical insignificance.

2) The BPS did not identify bar results by individual states, but only by regional groupings of states. Yet bar exams, and the proportion of takers passing, vary widely across states and within regions. In his original analysis, Sander found that this grouping did not bias overall measurement of the black-white gap in bar passage rates (Sander 2004). But there is no question that the lack of individual-state data substantially adds to the indeterminacy of analyses with BPS data.

3) As discussed earlier, the BPS data does not allow a researcher to identify individual schools, but only one of six sizeable and often heterogeneous "clusters". This greatly and intrinsically limits one's ability to control for a variety of school effects.

4) The BPS covered the cohort of students who completed law school in 1994, but, as noted above, bar exams have generally become more difficult since then (Merritt, 2001; Jones 2006) and outcomes for blacks appear to have significantly worsened (Sander, 2006b). Studies based on more recent data are therefore much needed.

The key to creating a richer dataset is to secure the cooperation of state bar commissions, which administer exams and maintain custody of records.

V. Our Proposal

Our proposal has two different aspects and goals. The first is to use California bar data to conduct four different types of tests of the mismatch effect – all of them of unprecedented rigor.

The second is to complete the first stage of building a multi-state dataset that will facilitate more comprehensive and generalizable analyses of issues related to law school preferences and possible mismatch effects within the next two to three years. We discuss these in turn.

A. Analyses of California Bar Data

For understandable reasons, state bar associations have traditionally been highly protective of data on bar scores.   At the same time, leaders of the bar are concerned with the large gaps in pass rates across racial lines, and many would like to increase the diversity of the legal profession. A well-credentialed, independent study of the minority bar passage problem, with strong safeguards against the disclosure of personal information, has an excellent chance of securing the cooperation of many state bars.

California is the ideal place to begin this work. California is the most populous state by far, and the State Bar of California (SBC) is second only to New York in membership and the number of test-takers each year. California has the most racially diverse bar in the country, with large numbers in each of the four major racial/ethnic groups (white, Asian, Hispanic and black); nonwhites make up nearly 40% of those sitting for the bar. The SBC has been the only bar in the country to regularly publish detailed results by race, and it has been the most progressive bar in its pursuit of in-depth research, including studies of the bar's internal validity and of ways in which the bar can better test skills involved in legal practice (one series of such studies led to the introduction of the so-called "performance test" section of the bar exam in the early 1980s, which has since been adopted in some form by most other states).

As we will detail below, much of this research has been done by Dr. Stephen Klein. Klein has conducted dozens of SBC studies over the past thirty years on issues ranging from internal validity to ultimate passage rates. He is not only intimately familiar with the data collection processes of SBC, but actually has most of its data in his possession and can vouch for its quality. We are in an advanced state of negotiations with the SBC to conduct the analyses described below, some of which involve the collection of supplemental data. Obviously, the decisions of NSF and SBC are, to a degree, mutually dependent.

It is useful to distinguish three SBC datasets. The first, which we will call SBC-1, consists of existing bar data for the exam cohorts from 1997 through 2006. These data include the raw and standardized scores of bar takers on the various components of the bar, the LSAT scores of most bar takers, their race and gender, and the law school they attended. The second dataset, which we will call SBC-2, consists of existing bar data collected by SBC for past validity studies. This dataset covers fewer years, but includes (in addition to the data in SBC-1) data on the law school grades and undergraduate grades of bar takers. The third dataset, which we will call SBC-3, would duplicate SBC-2 for two recent cohorts of California bar takers (those taking the bar in 2004 and 2005), and would include for nearly all bar takers their detailed bar scores, law school attended, race and gender, law school GPAs, undergraduate GPAs, undergraduate college LSAT scores. Only SBC-3 requires significant new data collection efforts. With this data, we will conduct the four analyses described below:

Analysis 1: An analysis of the impact of Proposition 209 on minority bar passage. Natural experiments are common in the social sciences when it is not feasible to conduct true experiments, such as when the focus is on the effects of new laws or policies. The adoption of Prop 209 in California in 1996 provides such a natural experiment. Prop 209 generally prohibited the use of race in state programs, including university admissions. Four California law schools

were affected: Boalt, UCLA, UC Davis, and UC Hastings. Boalt and UCLA are both elite schools that used preferences aggressively prior to the passage of Prop 209. UC Davis used preferences somewhat less aggressively, and UC Hastings still less. The differences in the numbers of minority students at these schools before and after the passage of Prop 209 suggests that the new law caused the schools to admit roughly fifty fewer blacks and fifty fewer Hispanics annually at these schools. As a practical matter, those denied admission to these UC schools could either attend less elite schools within California, attend Stanford (a private elite school in California), or attend elite public or private schools out-of-state. To the extent that they chose to remain in-state (and did not go to Stanford), they would have presumably attended a law school where their admissions credentials were more like those of their classmates.

Analysis 1 will begin by comparing the pre- and post-209 admission credentials of minority students at the various UC schools to identify the approximate the cohorts of students (within a credentials band) who were admitted to particular UC schools before Prop 209, but not afterwards. We will determine the bar scores and pass rates of these cohorts coming out of the UC schools during the three years before Prop 209. We will then identify which of the students in comparable post-Prop 209 cohorts ended up at less elite schools, and will examine the bar scores and pass rates of those cohorts. This provides a simple, direct test of the mismatch hypothesis, which predicts that the post-Prop 209 minority cohorts, to the extent they end up at less elite law schools, will have stronger bar performance than the comparable pre-Prop 209 cohorts.

It is also very plausible that the displacement of one hundred or so minority students by Prop 209 each year produced a more general cascade effect among California schools. For example, if the University of Southern California (USC), which usually loses its best applicants to UCLA and Boalt, in 1997 faced an upsurge of strong minority applicants, it probably used smaller preferences and rejected some of the lower-credentialed minority candidates it normally admits. These students then faced a similar choice of leaving the state or attending a lower-ranked school. This sets up a variety of related natural experiments.

The California bar data tracks both in-state and out-of-state applicants; each year, there are several thousand first-time bar takers from California law schools and about 1500 candidates from out-of-state schools. (Over the three years before and after Prop 209, about 1200 in-state blacks and 600 out-of-state blacks took the bar exam, along with a slightly higher number of Hispanics.) Out-of-state pools should be essentially unaffected by Prop 209. We thus have a rich set of comparison populations: minority cohorts in-state before and after Prop 209, and minority cohorts out-of-state before and after Prop 209. This sets up several tests of hypotheses:

(1) Proposition 209 went into effect for law schools in the matriculation year 1997 – students who would normally graduate and take the bar in 2000. We predict that for students taking the California bar from out-of-state law schools, there should be no change in bar scores before or after the passage of Prop 209 in relation to their LSAT scores.

(2) We predict that for in-state blacks and Hispanics, we will observe the following beginning in 2000: (a) a slight overall decline in the average LSAT of in-state blacks and Hispanics taking the bar; (b) a larger decline in the LSAT gap between individual blacks and the median LSAT of students at their law school, and between individual Hispanics and the median LSAT of students at their law school; (c) better overall performance in bar scores among in-state blacks and Hispanics relative to similar out-of-state blacks (controlling for LSAT); (d) higher bar

scores for black and Hispanic students, controlling for LSAT, when that student has a smaller credential gap with classmates.

All of the analyses described above uses data already on hand (SBC-1). The approach we describe nicely avoids the selection bias problems so pervasive in mismatch studies because the decline in credential differences is an average decline observed for an overall population, not an inferred individual decline based on a smaller test score gap. In other words, we have more confidence that smaller observed credential gaps between blacks and Hispanics vs. their classmates in California after Prop 209 are in fact the result of smaller preferences, compared with the pre-209 and the out-of-state populations.

Analysis 2: Cross-Section Analysis of Bar Score Outcomes using "Case Control" Method. A second basic approach to examining the impact of preferences is through the direct comparison of students with differing levels of credential disparities with their classmates. The straightforward hypothesis is that, other things being equal, we will observe worse bar scores for students (regardless of race) who have a large credentials gap with their classmates than when the gap is smaller.

In Analysis 2, we will construct a logit model to predict a student's likelihood of passing the bar exam based on that student's admission credentials (i.e., LSAT score and UGPA, with the later variable adjusted for undergraduate school attended). We will then form pairs of students where both members of the pair had comparable predicted probabilities of passing, but one graduated from a law school where his/her classmates had similar admissions credentials while the other member of the pair graduated from a law school where his/her classmates had a substantially different "credential gap" (i.e., much better or worse). We will then examine whether the difference in credential gap is related to differences in bar scores between members of a pair. This method of analysis corresponds to the type of "case control" study Klein, Freedman, and their colleagues used in their recent study of possible racial bias in the federal death penalty (2006).

There is some room for selection bias in this approach; in comparing a pair of students with similar entering credentials, is the one who is at the more elite institution (and thus potentially more mismatched) really comparable to the other student, or is he at the more elite institution because of some unobserved advantage? However, the bias is conservative since it cuts against a finding of mismatch effects. More to the point, we use three credential factors to compare student credentials: LSAT, undergraduate GPA, and undergraduate college. This third factor – college quality – has probably been the most important unobserved biasing factor in earlier analyses, so we believe selection bias will be quite small. And, of course, this test will be far more powerful than earlier tests in three other respects: the direct (rather than inferred) measure of credential gaps for each student, the use of actual bar scores (rather than pass/fail outcomes), and the elimination of confounding effects from using multiple bars.

Analysis 3: Cross-Sectional Analysis of Bar Score Outcomes with Regression Techniques. This analysis will be similar to Analysis 2 in all respects except for methodology. Analysis 2 uses a case-control method to pair comparable students. Analysis 3 uses standard regression techniques to compare all students in the population for whom we have complete data. The dependent variable in this model is a bar-takers' exam score. The independent variables include the student's credentials, gender, race, year, and the measured credentials gap between that student and his classmates (i.e., a measure of mismatch).

Analyses 2 and 3 will use two databases, SBC-2 and SBC-3. As noted earlier, SBC-2 includes detailed data on student credentials gathered by the Bar for earlier validity studies. SBC-3 will gather comparable data for the 2004 and 2005 bar-taking cohorts. In this process, we will enlist the cooperation of California law schools (there are nineteen ABA-accredited law schools in the state, and a number of unaccredited schools), using the confidentiality methods described in Part VIII, to gather information on student credentials and law school performance. We estimate these data will contain complete information on about eight hundred blacks, twelve hundred Hispanics, two thousand Asians, and several thousand whites. These sample sizes are sufficient to allow reliable analyses within racial groups, and using racial controls in cross-racial analyses. The sample sizes are also more than sufficient to allow exploration of general and within-race gender effects, as discussed below.

Analysis 4: Descriptive Analyses of Preference Patterns. The prior three analyses focus on measuring possible mismatch effects among law graduates. This analysis examines instead a range of other issues relevant to a discussion of admissions preferences in legal education. These include: (a) the current extent of racial admissions preferences by California law schools; (b) the predictiveness of law school admissions criteria in eventual bar performance; (c) the relation between admissions criteria and performance in law school, including the question of whether blacks and Hispanics over- or under-perform standard admissions criteria. Each of these issues is important in understanding the scope and remedies for mismatch effects.

Capturing subtle effects and avoiding bias. The central goal of our analyses is to study mismatch effects in ways that minimize selection bias and avoid false positives (i.e., finding mismatch effects where none exist). Our research design effectively addresses the biggest problem – bias caused by unobserved credentials. But we have also taken into account several other factors related to bar exams.

Gender. Men tend to slightly outperform women on the multiple-choice portions of the bar exam; women do better on the essay portions, but the differences are small. We will have data on the gender of all participants, and will use gender not only as a general control variable, but as an interaction term (with race) to explore possible variations in the role gender plays across racial groups.

Selection into the bar exam. In theory, our analysis could be biased if a large number of law graduates did not take the bar exam, and if the non-bar-takers were concentrated at particular types of schools. In fact, the BPS found that 94% of all law graduates take a bar exam, with little variation across racial lines (e.g., 93% of black law graduates take the bar). The grades and school eliteness of those taking the bar are distributed much the same as those not taking a bar exam. We are therefore not concerned that our focus on bar results will produce a skewed picture of the preparation level of different groups.

Another possible source of selection bias is the higher attrition rate *during* law school of students at less-elite schools. Since our data focuses on law graduates, the attrition would be a concern if a larger proportion of the students we wish to compare were "weeded out" before graduation at less-elite schools than at more elite schools, since that would skew the sample. In fact, the higher attrition at less-elite schools is driven by their admission of lower-credentialed students. When one controls for credentials, graduation outcomes are similar or arguably higher for students at less-elite schools (Sander, 2004, 2005; Rothstein & Yoon, 2006). Thus, we do not

think this is a significant form of bias, though it is an issue we will monitor closely in our research and one we will try to address further by securing information on cohorts starting law school.

Socioeconomic background. Our variables will not include a measure of class, and we do not believe there is a feasible way of including such a measure in an analysis that makes use of many past years of bar data. We do not believe this is a meaningful omission. Class does correlate strongly with tests such as the SAT and with success in achieving a bachelor's degree or pursuing graduate education. But a number of studies have found that class is not an important predictor of either law school performance or bar performance (Wightman, 1998, Klein, 2004, Sander, 1997). And since students at lower-ranked law schools tend to have somewhat lower SES backgrounds than students at higher-ranked law schools, any bias introduced by an SES influence on bar performance would tend to be conservative with respect to our mismatch hypotheses.

Bar preparation courses. Some of the data we will collect includes information on the bar preparation activities of respondents, but most will not. Is this an important omission? We think not, for a few reasons. First, preparation activities are quite uniform. According to BPS data, 96% of respondents (and 93% of black respondents) who took a bar exam reported they had taken a commercial bar review course. Second, (as with SES), bar prep activities have a tiny impact on bar performance, compared with student credentials and law school performance. Klein's (2004) study of performance on the Texas bar found that, for the nine law schools studied, LSAT and law school GPA predicted, on average, about 44% of the variation in bar scores. Adding four bar preparation variables increased the average R^2 of these models from 44% to 47%, and most of these added variables were not statistically significant. Third, as with SES, lower-tier graduates are somewhat more disadvantaged in bar preparation than higher-tier graduates. In the Texas study, graduates of the five lowest-ranking law schools were about twice as likely to not take a commercial bar lecture course and fifty percent more likely to work part-time while preparing for the bar, compared to graduates of the four high-ranking law schools. Thus, any residual impact of these variables is conservative with respect to the hypotheses we are testing.

B. The Development of a National Database on Legal Education Outcomes

The analyses just described can shed enormous light on the questions of peer effects in learning and mismatch effects in preferential admissions. But two additional types of data would round out the picture. Data from additional states would allow one to test the external validity of the California results. And data on law school applicants and admission patterns would make it possible to use even better selection-bias controls in the analysis of outcomes, by finding pairs of students admitted to the same pair of law schools but chose to attend schools of differing eliteness (the Dale-Krueger method described in Part I, above).

We believe the key to building a better database is to create a foundation on which other institutions can build. We have had discussions with several of the major bars around the country. Like the State Bar of California, they are aware of the mismatch issue, interested in getting to the bottom of the sources of racial disparities in bar outcomes, and desirous of better understanding dynamics within their own states. A natural outgrowth of our work with the California data would be the sharing of findings and exploration of how complementary studies

can be done with data from other states. This process is already underway, and we have already made significant strides in creating a national network of scholars interested in the issues and the data.

Another key institutional player in this process is the Law School Admissions Council (LSAC). LSAC administers the entrance examination for law schools (the LSAT) and provides services that facilitate the admissions processes. It generally tracks the schools to which students apply, which ones accept them, and where they initially attend. It also widely collects data on first-year performance in law school for the purpose of conducting validity studies that help schools determine how much weight they should assign to the LSAT (relative to undergraduate grades) in the admissions process.   Unlike the state bars, which are generally governed either by the state judicial system or the associated lawyers of the state, LSAC is governed by the accredited law schools. Given the intensely political debate about preferences within law schools, it is not surprising that LSAC has been cautious about undertaking its own research on mismatch issues. Once again, however, we believe that the emergence of a multi-state, contemporary database of bar outcomes, and a broad network of involved scholars, will create a very powerful inducement for LSAC to cooperate.

VI. Intellectual merit and broader impact

Our study has strong **intellectual merit**. The "mismatch debate" has been one of the most active debates in legal academia in recent memory. Many economists have become quite interested in recent years in sorting processes in education, peer effects, and the determinants of human capital gains in education – all questions directly implicated in this project. Our proposal brings together two economists, two law professors, and a leading psychometrician, and creates a database far superior in many respects to any developed so far on these issues. The results of this study will substantially advance several discourses.

Our study will also have very **substantial broader impact**. The debate over racial preferences in higher education has been one of the most important and long-lasting in modern America. Legal education – as reflected in the *DeFunis*, *Hopwood* and *Grutter* court decisions – has always been near the heart of the debate. Our study is likely to receive national attention and to help make more informed, and more effective, efforts within the legal academy to promote racial justice and inclusion. If the study finds significant evidence for mismatch effects, this may encourage efforts by law schools to provide better academic support, better disclosure to admitted students about the tradeoffs involved in particular school choices, and more attention generally to the "peer structure" of the law school environment. Whatever the study's findings, they will provide empirical common ground in an area rife with contention.

VII. Project Team

The five principal investigators on this proposal are uniquely well-positioned to carry out this study. Dr. Richard Sander is an economist and law professor at UCLA, and probably the leading author in the country on legal education issues over the past decade. As noted, Sander's analysis of law school mismatch effects integrated a wide range of previously unpublished data and has generated an extraordinarily active debate in legal academia.

Dr. Stephen Klein is one of the nation's leading psychometricians and the unquestioned dean of research in bar studies. As a Senior Research Scientist at RAND over the past thirty years, Klein conducted pioneering research on methods of improving the bar's measurement of lawyering skills and has been a key player in the recent development of the Collegiate Learning Assessment Project, which uses innovative testing methods to assess learning during college, and has been adopted by over two hundred colleges nationwide. Dr. Klein did a very early (1988) study of how law school peer effects might affect bar performance; this preliminary research was skeptical of mismatch effects, though it admittedly had quite limited controls for selection bias.

William Henderson is an associate professor of law at Indiana University. Henderson brings several important things to the project. His work on the LSAT (2004) is widely thought to be the most important and considered recent critique of standardized testing in law school admissions. Henderson thus brings a healthy skepticism about the meaning and interpretation of test results. Henderson is a leader among the younger generation of scholars studying the legal profession, and he has played a prominent national role in encouraging balanced and dispassionate discussion of Sander's research on law school mismatch effects.

Dr. E. Douglass Williams is Chair of the Economics Department at Sewanee University (aka The University of the South). Williams is a cautious, careful labor economist whose interest in legal education goes back to his PhD dissertation which included a chapter on the role of law schools in the rapid growth of the lawyer population during the 1970s.   Williams knows well the economics literatures on peer effects in learning, and human capital acquisition in higher education, and he is a skilled econometrician.

Vik Amar is Professor of Law at the University of California, Hastings School of Law. He is best known as a scholar of constitutional law, but he has written widely on subjects relevant to this research, including bias in testing, retooling law school admissions, the effects and constitutionality of Proposition 209, and the mismatch effect. Amar has worked for many years at Hastings to improve the success of the school's graduates on the bar.

The two other staff in the project are Dr. Roger Bolus and Dr. Robert Sockloskie.   Bolus received his Ph.D. from UCLA in Research Methods, Measurement and Evaluation. He is founder and senior partner of Research Solutions Group, a consulting firm that specializes in providing psychometric, statistical analysis, and database management services to large scale research programs; and he has coauthored many analyses of bar data with Dr. Klein.   Dr. Sockloskie has a Ph.D. in psychology from UCLA, has worked at RAND, and has worked for the past eight years as a data analyst for the Empirical Research Group at UCLA's School of Law. Dr. Sockloskie is a highly-skilled methodologist and statistical analyst.

VIII. Past NSF Support

Dr. Sander was co-PI on NSF grant 0115521. *After the JD, Phase I* was funded in 2001; the NSF grant of $166,000 covered approximately ten percent of the total cost of this longitudinal study of nearly five thousand lawyers who completed law school in 1999 and 2000. An initial report on the project was published in September 2004, entitled *After the JD: First Results of a National Study of Legal Careers*. Sander has used data from the study in three recent articles (Sander 2004, 2005, 2006). Part II of the study, which will re-survey participants

in the seventh-to-eighth year of their careers, received support in 2006 from NSF's Law and Social Science Program (grant 0550605). Sander is not a co-PI on Part II.

Dr. Klein was a PI on NSF grant ESI-99-86612, which ran from 1999 to 2005 and was entitled *Mosaic II: Longitudinal Investigation of the Effects of Systemic Reforms on Student Achievement*. Mosaic II was a six-year longitudinal research project sponsored by the NSF and co-directed by Dr. Klein and Dr. Laura Hamilton. It was a continuation of RAND's (and Dr. Klein's and Hamilton's) research on the relationship between teacher instructional practices and student achievement in mathematics and science. Results from their earlier NSF studies (grant ESI-96-15809) suggested that exposure to reform-based instruction was only weakly related to student achievement in math and science. Mosaic II allowed for a more in-depth examination of this relationship, including more complete measures of classroom practices, the use of NSF-developed curriculum materials, and exposure to various staff development opportunities. The results of this research culminated in several reports and journal articles, including Vi-Nhuan Le et al (2006), Lockwood et al (in press), Stecher et al (2006), and Hardison et al (under review).

IX. Confidentiality procedures and Human Subjects Issues

This project does not involve the collection of any new data from individuals, but simply the assemblage of existing data. In two of the datasets we described in Part V, individual identifiers have already been removed; in the third (SBC-3), which involves adding data collected by law schools to data held by the State Bar, we will create a codebook, held only by Dr. Klein, that will generate two random identifiers for each participant – one identifier labeling bar data and the other labeling data collected from schools. The project team will not observe any personally-identified data, and any shared versions of the dataset will suppress data on cohorts at particularly institutions with fewer than five members.

As we gain approval from California and other state bars granting access to specific sets of data, we will secure appropriate approvals for the analysis of that data from the UCLA Human Subjects Review Committee.

X. Budget, Partial Funding and Conditional Funding

Our budget is quite modest relative to the analyses we propose and their potential academic and policy payoff. The budget is low because (a) four of the five PIs will not receive any compensation and (b) the bulk of the data we will analyze is already gathered and cleaned. As noted earlier, we are in the process of securing approval for analysis of the California data, and we are in various stages of negotiation with several other states. We would expect NSF to condition specific phases of project funding on the execution of agreements securing various databases for our use.