November 6, 2007

Memorandum:

To: Board of Governors, State Bar of California
From: Richard Lempert
RE: Data Request by Professor Richard Sander, et al.

Overview

I am Richard Lempert, the Eric Stein Distinguished University Professor of Law and Sociology at the University of Michigan. I have written several empirical pieces dealing with affirmative action, including work responding to earlier publications by Professor Richard Sander. I have also followed from afar the efforts of Professor Sander and a team he has assembled to access confidential California bar test data to test his "mismatch hypothesis." One law professor concerned about the likely quality and thrust of this research asked if I would fly to California to present my views on the data request. I am, however, otherwise occupied on the day of the meeting, but I said I would read relevant material and put my views in writing. I apologize in advance for the length of this memorandum. I found that there is much more to say about the merits of the Sander *et al* proposal (hereinafter "Sander" or "the team," with "Professor Sander" used when I wish to refer only to him) than I had anticipated.

For reasons given below I believe the California Bar's Board of Governors finds itself in a position it should not and does not deserve to be in as it reviews the appeal by Sander from the Bar Examiners' decision to deny the team's data request. Indeed, I cannot imagine the Board of Governors reversing its subcommittee's decision except on political grounds. By this I do not mean that I believe the Board is hostile to affirmative action and hopes Sander's proposed research, like Prof. Sander's past research, will lead some to call affirmative action into question. Rather Sander, through an op ed column in the LA Times and otherwise, has sought, with some success (witness Stuart Taylor's column in the National Journal), to portray the Bar Examiners' decision to deny Sander's data access request as a decision that was largely motivated not by privacy, scientific or legal considerations but rather because the Examiners wished to avoid controversy or, even worse, because they wanted to take the "politically correct" action. Today, ironically, charges of "political correctness" allow for only one "correct" response: refuting the charge by acceding to the demands of those who claim they have been thwarted by political correctness. Just as the Bar Examiners' would have been wrong to base its decision on fear of controversy or in what the results of research might show, so

the Board of Governors would be wrong to decide Sander's appeal because its members have been lobbied by interested parties or are concerned with the political context of how their decision will be viewed. Rather the Board should evaluate the Bar Examiners' decision on the merits, and if it believes the subcommittee's grounds do not sustain its decision, the Board should address further issues relevant to the data request that the subcommittee felt it did not have to reach. My expectation is that if the Board does this, it will conclude that the proposal in its current form is inadequate to justify the release of bar exam data. Indeed, I think the Board should conclude that absent certification of IRB approval, this proposal is not even ripe for consideration.

Let me note, at the outset, that I am a strong supporter of empirical work, and, in particular, a believer in work relevant to policy. Much of my own research has been of this sort. Moreover, I do not think social science research should be hampered or suppressed because some groups, even powerful pressure groups, believe the results of well-conducted research will be uncongenial to their preferred policy preferences. With respect to law school affirmative action, I believe sound empirical work can be a win-win proposition whatever it reveals. No one gains when students admitted to law school through affirmative action fail to benefit from their education because they do not graduate and pass a bar. If we can better understand why some students, including members of certain racial groups, have special difficulties in graduating law school and passing the bar, then we might be able to improve the situation, either by better advising students about paths that make sense for them to take or by changing how we admit, educate and test law students.

Empirical research need not be perfectly done or answer all questions to make valuable incremental contributions. But unsound research, and research that does not meet high standards of scientific integrity, can harm by misleading both the public and policy makers about what available evidence shows. Moreover, even sound and valuable empirical research can do harm if it adversely affects human subjects. For the latter reason, Universities and other research organizations require Institutional Review Board (IRB) approval for work involving human subjects done by people in their employ. In deciding whether to provide Sander with the data the team seeks, the Bar should consider both human subjects issues and the likely quality and value of the proposed research. In addition, given the serious questions raised about Prof. Sander's data analyses and interpretations in his earlier work on affirmative action, the Bar should not provide the Sander team with data access unless it is prepared to do the same for other scholars who would wish to re-examine Sander's analysis and results. The Bar Governors should recognize that, for reasons that have nothing to do with Dr. Steven Klein's well-recognized integrity and skills, effective checking cannot occur if those who would check Sander's work must rely on Dr. Klein and/or Dr. Bolus (whom I do not know, even by reputation) to do the actual data analysis.

Human Subjects Protection

I am unfamiliar with the Bar's rules and policies on data access for bar-test-relevant or more extensive empirical investigations, and I am also unfamiliar with state or federal laws that may constrain the Bar's discretion. Other law professors can help the Bar more on these issues than I can, though my understanding is that there are issues under both federal and California privacy laws that may proscribe the Bar from releasing the data sought. As a researcher, however, I hope there is leeway for scholars to access recent data for well-conceived investigations that go beyond simple test validation research. If, however, there is no policy or legal leeway to allow this, that is the end of the matter, and I should stop writing and you can stop reading now. I continue on the assumption that there are circumstances in which the Bar data can be used for research of the kind Sander proposes.

To release its data the Bar must be confident that conditions for the ethical use of human subjects data are met. This involves meeting general social science standards for human subjects protection as well as satisfying any special concerns the Bar has. The Bar, however, has no special expertise on general social science standards for the protection of human subjects, and even if it did, its approval would not be sufficient to allow the research to proceed. Rather the determination must and should be made by an IRB. In the case of the Sander proposal, it is, as Sander acknowledged in his NSF proposal, the UCLA IRB that needs to approve the research. Because the need for human subjects approval can lead to changes in research proposals, the proposal the Board of Governors has before it may not describe the research as it will be carried out.

For this reason the Bar should not now be facing the need to respond to the Sander data request or to an appeal from its Bar Examiners' decision. Rather the Bar should ask Sander to secure human subjects approval from the UCLA IRB and then return to the Bar with an approved proposal. In this connection the Bar may wish to note that Sander's two unsuccessful proposals to the National Science Foundation say different things about IRB approval. The earlier proposal says, "We are *in the process of* securing approvals from the UCLA Human Subjects Review Committee." (B-15) The later proposal says, "*As we gain approvals* 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." (B-15) [Emphasis added.] This sequence is strange. Unless the UCLA IRB had problems with the Sander proposal or Sander for some reason decided to halt the approval process, it should have been possible to secure an IRB decision by the time of the second NSF submission. The Board may wish to inquire whether in the original process of securing IRB approval issues regarding human subjects were raised.

This fact that human subjects approval could have been obtained by now but apparently has not been is all the more reason why, if the Bar's Board of Governors is willing to consider reversing the Bar Examiner's decision, it nevertheless should not act at this time. Rather, to use lawyer's language, it should remand the proposal as not ripe. The Bar's need to confront the data access issue can be postponed until Sander's proposal is approved by the UCLA IRB. If the issue arises, it may arise in the context of a proposal that looks different from the proposal before the Bar now, and it may not arise at all. Moreover, in deciding Sander's request, the Board might be aided by knowing the proposal has secured IRB approval. In particular there are at last three human subjects issues the Board will want to think about. To aid it in its decision making, the Board could ask Sander in his request to the UCLA IRB to highlight them for special consideration.

The first is the issue of informed consent. The letter Sander sent (with Amar as first signer) to the Bar following the Bar Examiners' rejection of the team's data request, sees the Bar Examiner's decision as inconsistent with the California's Bar's cooperation with the LSAC Bar Passage Study (BPS). All participants in the BPS had, however, consented to the release of law school and bar passage data. I cannot speak for the State Bar of California, but I expect such consent figured importantly in the Bar's data release decision.

Informed consent issues can be difficult ones, and the UCLA IRB has, no doubt, considerable experience dealing with them. The Bar Governance Board could make good use of the benefit of their thinking, although it may have concerns beyond the issues that would concern an IRB.

The second issue for IRB review is whether the Sander proposal guarantees adequate protection for the institutional anonymity of law schools and the individual anonymity of bar applicants. The proposal to the bar specifies that it will not break down the data by category if a category contains four or fewer individuals. An IRB may well regard this as inadequate human subjects' protection, and even if the IRB does not, the Bar may. As Sander noted in his NSF proposal, a weakness in the BPS data set is that it lumps the nation's law schools together in only six tiers with ambiguous boundaries. Those who designed the BPS researchers did not want a less tractable design, but they feared finer breakdowns would endanger the privacy interests of schools and the students who attended them. It is these interests that Sander's research may threaten. If references to racial grouping involving as few as four students can be made, a statement like, "75% of the black graduates at an elite private California Law School who took the bar for the first time in 2005 passed with only a point to spare" is in theory permissible. But the only elite private law school in California is Stanford, and most 2005 Stanford graduates, as well as curious others, might be able to identify the black students referred to, tarring them all with a 75% "barely passed" brush. Indeed, aggregate test score disclosure can do more to invade privacy of a bar applicant who is part of a small cohort than a statement that only a certain percentage of the cohort passed the bar. This is because apart from the disclosure, no test score information would be available to anyone, but information about who has passed the bar is public, and interested persons

who know the composition of a small group that took the bar can infer that those not listed as passing most likely failed.

An IRB would want more insurance of confidentiality protection than a researcher's promise to wisely use discretion in making references to small groups. This caution resonates in the light of a recent op ed column that Prof. Sander and a fellow team member placed in the *LA Times*, seemingly in an effort to make the Bar Examiner's rejection of their request a public issue. The column says that graduates in the class of 2005 from "one selective California law school" who benefitted from strong "preferences" fail to pass the California bar at ten times the rate of those who received no preferences. Since in a later web posting this selective school is labeled "elite," it is almost certainly Professor Sander's school, UCLA. That is the only elite California school whose bar pass data he is likely to have been able to access unless someone who knew the data for another school violated an ethical obligation of confidentiality. Moreover, although the strong preference group is not defined in racial terms, the implication in the column, confirmed by Sander's web elaboration, is that these were mainly black and Hispanic students. Given Prop. 209 there cannot be large numbers of black and Hispanic students who graduated in from UCLA in 2005, and then took the California Bar. Hence UCLA graduates of that year as well as some law firm and other colleagues of 2005 UCLA graduates are likely to know, or think they know, who these "strong preference" students were.

If statements in op ed columns required IRB approval, I doubt if this one could have received it. Moreover, not only is this statement a potential infringement on privacy interests, but it is poor social science. Rates based on a small number of cases at a particular point in time are often unstable. A more reliable, although no doubt less striking, estimate of student difficulty would have combined data from a number of years. As a more aggregated estimate would have included more students, it would also have been more protective of student privacy.

In one of the team's follow-up letters to the Bar, Sander appears to recognize that there is an issue here and indicates that in no case would a single school constitute a category. This is more protective of privacy than the fewer than five in a category commitment made in the proposal, but it still would allow statements pertaining to few students and affect institutional interests. Any reference to an elite California public law schools could, for example, refer only to UCLA and Boalt Hall while references to non-elite public California law schools would be read as "Davis and Hastings." Moreover, the number of black students from elite out of-state-law schools who take the California Bar may, even when totaled, be small.

The Sander team's letter goes on to suggest that, strictly speaking, even breakdowns into larger groupings may not be necessary to study the implications of mismatch, but it make no guarantees the research will ignore school subcategories. Nor should it at this stage of the design process, for there are scientific costs to avoiding subcategorization. The trade-offs involved raise the kinds of issues that can only be resolved with external expert guidance in the IRB review process. That, not the State Bar of California, is the appropriate venue for thinking through questions of this sort.

Finally, the IRB should be asked whether it sees any human subjects issue in the likelihood that all minority bar takers might suffer from the disclosure of and subsequent publicity given to minority group California Bar Exam scores. Even though some minority bar takers, no doubt do very well on the bar, given that minorities fail the bar at higher rates than whites and as a group have lower LSAT scores than whites, it is predictable that non-Asian minorities will have, on average, lower bar pass scores than whites. Yet minorities who pass the California Bar pass one of the nation's toughest bars, and they perform at a level which the California Bar sees as providing a substantial guarantee of competence for legal practice. Revealing gross test statistics by race may nonetheless, in the mind of the public, degrade the apparent or relative competence of all those in a group identified as having a low average bar pass score. This is true even though numerically, rather than on average, there may be more whites and Asians than blacks, Hispanics and Native Americans combined among those with the lowest bar passage scores. Looking at 1985 and 1986 bar takers, for example, Steven Klein found there were about 4 times as many Anglo students with LSAT scores in the bottom quartile as there were black and Hispanic students together, yet on average Anglos had substantially higher LSAT scores than blacks or Hispanics. A person with the (admittedly silly) concern that too many blacks and Hispanics with low LSATs were becoming lawyers, would most likely be blithely ignorant of the fact that most of those with low LSAT scores who became lawyers were white. It is for reasons like this that, unless carefully presented, even valid social science research results can distort public perceptions and lead to poorly informed policies.

Clearly the bar must be concerned about the possibility of degrading the value of passing the bar for lawyers who belong to certain racial and ethnic groups and for lawyers of all races who followed evening programs or attended non-ABA accredited or otherwise low ranking law schools. Indeed human subjects concerns might lead an IRB to insist that the proposal be rewritten so that race and school and program type would not be associated in any publication with bar pass scores. This is feasible because if the study is designed to test the mismatch hypothesis rather than to examine how blacks and Hispanics do on the bar, only degree of mismatch must necessarily be linked to observed outcomes. Race becomes relevant only if mismatch effects are found and differ with ethnicity. This would suggest that something other than sheer mismatch was responsible for part or all of identified mismatch. Even in this case, however, the average scores that different racial groups received on the bar might not be an important datum. School and program type may be even less likely to matter.

Before leaving this topic, there are two final points to be made. First, Sander could make a formal argument to the UCLA IRB that this research, since it is based on an already existing data source with no names attached, should be exempt from IRB approval requirements. The Bar should make clear that it will not accept an exemption determination as IRB approval. Rather Sander should request that the IRB examine the proposal in light of the above issues and such other issues relating to human subjects protection as the Bar or IRB might identify. Second, Sander has responded to particular criticisms of the team's proposed design on a piecemeal basis. Even if the Board does not wish to postpone its decision pending IRB approval, it should not act until it receives a coherent proposal that assimilates in one place Sander's planned ways of dealing with what that team acknowledges to be legitimate concerns.

Design and Quality Issues

It might seem that IRB review is a significant hurdle for the proposed research, and I believe it is, but it is not an impossible one to surmount. An IRB may view the proposal's human subject protection issues differently than I do, and even if IRB members share my concerns, the proposal may be redesigned to alleviate these concerns. Moreover, even legitimate human subjects concerns are not dispositive in the IRB review process. Rather IRBs balance the value of research against its potential for harm to human subjects. Research that poses some human subjects risks may nonetheless be approved if its likely scientific and social contributions are substantial. This requires, however, an assessment of the likely value of the research.

Similarly the Bar's Board of Governors in deciding whether to release the data in its possession should decide whether the available data and the proposed research design are likely to allow contributions significant enough to justify the release of the data. If the research plan and available data will not allow as powerful tests of the mismatch hypothesis as Sander claims, then it may muddy rather than clarify the waters that swirl around mismatch theory and the wisdom of affirmative action. If so, the case for releasing the data substantially weakens. Thus, if bar policy, federal and state privacy laws, and university human subject protection protocols do not preclude the requested data access, then the Bar, if it wishes to act responsibly, must make an effort to assess the likely quality of the proposal before it and its likely contribution to the issues it purports to be able to clarify. This effort may be postponed and made unnecessary if the Board of Governors require UCLA IRB approval before they decide the data access issue, but if the Board of governors is satisfied that the proposal raises too many quality concerns to justify cooperation regardless of how other issues are resolved, it could save Sander and the UCLA IRB some work by reaching that conclusion sooner rather than later. For this reason I will discuss some measurement and design concerns that I have with the proposal as it now stands.

As with the human subjects issues, a decision by the Bar on the likely scientific value of the proposed research may be premature at this time. Neither the Bar Examiners, who first evaluated this proposal, nor members of the Board of Governors, are professionally well equipped to evaluate the design and measurement issues the proposal raises. The Bar could defer to the judgment of the National Science Foundation's Law and Social Science Program which decided not to fund this proposal and reject the proposal for not surmounting the NSF's peer review process, or it could appoint a small committee of social scientists to advise it on the likely fruitfulness of the proposed research. The California universities have numerous outstanding social science methodologists with no known dog in the affirmative action or mismatch hypotheses fights (e.g. Berkeley's Daniel Rubinfeld and UCLA's Richard Berk) who could help the Bar assess the proposal's quality. A committee of such experts might decide that data or methodological issues pose insurmountable obstacles to achieving Sander's objectives or, as importantly, it might suggest design and methods modifications that strengthen the proposal. It could, of course, also decide that the proposal merits the Bar's cooperation as written, but for reasons I indicate below, I doubt if this would be its conclusion.

A core problem with the proposed research is how it operationalizes (measures) law school learning. This is the issue on which the Bar is perhaps least in need of outside expert advice. The proposed research turns on the *assumptions* that California bar test scores are a *good* measure of how much was learned in law school and (apparently) that differences in test scores, even small ones so long as they are statistically significant, reliably reflect *meaningful* differences in learning. These assumptions are problematic.

First, some bar test takers may not have taken courses in law school on some of the subjects that appear on the bar exam. When I took the bar, for example, I was tested on commercial transactions, a subject I had never studied and on legal ethics, which was not then a required "for credit" course at my school. Another student, who learned as much as I did from his coursework and who had taken commercial transactions and legal ethics courses would have appeared to have learned more in law school than I by Sander's proposed measure. But had the bar exam instead tested on conflicts of law and administrative law, which I did take, I might have appeared to have learned more. With large numbers of subjects such variation in course selection would not be a concern if it were random with respect to the degree of mismatch or what the bar exam covered.

"Mismatch" will, however be greater in more selective than in less selective law schools, and variation across schools in the selection of bar relevant coursework is unlikely to be random or otherwise ignorable. Less selective law schools are typically more bar-oriented in their course offerings than more selective law schools, and their students are reputedly more bar-oriented in their course selection. Comparing students from a highly selective school with similarly credentialed students from a less selective school would therefore be likely to compare students who had taken fewer bar relevant courses with students who had taken more bar relevant ones. Differences in bar test

scores could not be safely attributed to degree of mismatch. A major weakness of the Sander proposal is that it does not recognize this possibly fatal weakness in its measure of law school learning and hence does nothing to address it.

A second problem with the assumption that bar test scores are a good measure of law school learning is that many of the subjects tested on the Bar students will have been studied two, three or more years earlier. Testing on material years after relevant coursework is unlikely to be a good measure of how well the material was learned in school. If it were, bar review courses might be out of business. But most bar takers take bar review courses, and a good portion of the variation in bar exam scores could reflect variation not in what was learned in law school but rather in the quality of an applicant's bar preparation and time spent studying for the bar. If students at less selective schools spend more time studying for the bar than similarly credentialed students at more selective law schools or if they otherwise are advantaged in bar preparation, degree of mismatch and the quality of bar preparation will be, perhaps hopelessly, confounded.

Sander recognizes this possibility and responds to it in the team's second NSF proposal. Sander told NSF that in the BPS 96% of respondents who took the bar, and 93% of black respondents who took the bar, said they took a commercial bar review course. I am unsure where this statistic came from, for I cannot reproduce it. When I looked at the BPS data for those who returned the BPS third follow-up questionnaire, which inquired about bar exam preparation, I found that about 82% of those responding who had taken the bar at least once said they had taken a commercial bar review course. I did get a figure very close to what Sander told the NSF (95%) when I looked at responses to the BPS second follow-up questionnaire, but the second follow-up only asked respondents how they *planned* to prepare for the bar. Because the bar was still in the future it could not inquire into how respondents actually prepared. The bar should ascertain the source of the figure Sander gave the NSF, and if the team chose to give NSF data on plans rather than data on experience the Bar should ask why.

The seriousness of the threat that differences in bar preparation pose for testing the mismatch hypothesis cannot be definitively ascertained from the BPS data because non-response to the third follow-up survey may be biased by bar plans. The available data suggest, however, that whether differential bar preparation is likely to be a serious threat to the validity of the bar score measure depends on whether Sander's matching on credentials controls for race, something which the NSF proposal seems *not* to contemplate. Although affirmative action eligible minorities on the one hand and whites and Asians on the other do not show consistent patterns in commercial bar preparation by school tier (except commercial preparation is highest among elite law school graduates in both groups) in every tier whites and Asians taken together are more likely to take a commercial course as part of their bar exam preparation than blacks, Hispanics and American Indians as a group. Of special concern is that among whites and Asians at schools in the two lowest tiers the proportion who report taking a commercial bar preparation course is higher than the proportion of affirmative action eligible minorities in any tier who report taking such courses. For many tiers, though not the elite tier, this difference is substantial. For example, the BPS data indicate that 86% of whites/Asians

in tier three schools took a commercial bar review course compared to 72% of minorities in the near elite tier, 73% in second tier public institutions and 67% in second tier private institutions. Since minorities in these tiers are likely to be closest in their index credentials to whites/Asians in tier 3 schools or historically black schools (where whites/Asians have an 89% commercial bar preparation rate) the danger that apparent mismatch will actually reflect differences in the quality of bar preparation is obvious. This differential preparation problem would most likely be ameliorated if matching were done within ethnic groups, but there are likely to be too few minorities with similar credentials in widely separated tiers to make such matching possible.

Sander also seeks to defuse concerns that differences in bar exam scores may reflect differences in bar preparation by citing research by Klein on the Texas bar. I have not seen the Klein paper and have had no access to the data, so I cannot directly assess the strength of Sander's arguments, but certain observations are possible. One claim is that after LSAT scores and law school grades are included in a model, 4 bar preparation variables add little to the explained variance (though it would appear that collectively they added a statistically significant amount.) Without knowing what the bar preparation variables were and their variance as well as the variance gained when law school grades were added to the model, it is not clear what to make of this.

The other claim is that any bias bar preparation introduces would run counter to the mismatch hypothesis because the Texas data show that in comparison to students in the highest ranking schools, students in the lowest ranking schools were about twice as likely to have *not* taken a commercial bar review course and about 50% more likely to work part time. The first part of this defense is troublesome for what it suggests about a possible willingness to obfuscate findings. "Twice as likely" may seem impressive, but when it is used to compare proportions of students *not* taking a commercial course it is almost completely uninformative. The statement would, for example, be true if 98% of students in the highest ranking schools took a commercial course and 96% of those in the lowest ranking schools did the same. Surely a better way of describing such a relationship would be to say that regardless of school ranking, the proportion of students taking the bar was almost the same. If the proportions taking the commercial bar prep course were 90% and 80% the statement would also be true, but the best way to describe the data would be to present these figures.

With respect to likelihood of working while studying for the Bar, the BPS data suggest the picture is not so clear. Consistent with Klein's data, graduates of elite tier law schools are less likely to be employed when studying for the bar than those in other tiers. But they are also less likely to put in more than 40 hours a week studying for the bar than those in other tiers, and considerably less likely (26% to 36%) than graduates of third tier law schools. Elite law school graduates are also less likely than students at lower tier law schools to start their bar preparation more than 8 weeks before the bar examination. There are relevant differences between other tiers as well. Some of these favor the mismatch hypothesis and others disfavor it . Overall, however, it appears from the BPS data that one cannot dismiss the possibility that bar preparation differences

might explain results that Sander would attribute to law school learning, and the study design does not seem to guard against such misattribution.

A third problem with the bar score measure concerns the meaning of differences. With large enough numbers of test takers, a difference of a single point on a bar test score could be statistically significant. Does this have any practical implications? When confronted with two people whose bar exam scores differed by a point, would we say that the one who scored higher learned more in law school than the other? Problems are compounded when group averages are used, for although the differences become more reliable, some members of the higher scoring group will have scored lower than some members of the lower scoring group and vice versa.

Even if there were score differences associated with mismatch that we were willing to attribute it to law school learning, how would we decide if they mattered? How large a difference would it take to conclude that a student at Stanford would have been better off attending a school in the middle of California's prestige hierarchy and foregoing the personal connections, job possibilities and mind-stretching but bar untested courses that Stanford offered? The Board of Governors would be well served if Sander offered it, in advance of looking at the data, some idea of the size of the bar score differences that the team expected to be associated with mismatch and how large differences would have to be to be regarded as meaningful.

Additional design problems are identified in a memorandum that William Kidder (who has been a coauthor of mine on several critiques of Sander's earlier work) provided the Bar Subcommittee when the Sander proposal was first considered. In his memorandum Kidder identified the proposal's most serious, and unfortunately potentially fatal, design problem: *only students who graduate and take the California bar can enter the sample*. The proposal most likely cannot deliver what it promises because it fails to account for those who begin law school but never take a California bar exam, either because they drop out of law school, graduate but never take a bar, or choose to take a bar elsewhere. The nature of this problem is well appreciated in other areas of educational research. An example from this sphere will clarify why this is a potentially serious and perhaps fatal problem.

Suppose, for example, that a researcher wants to know whether a pre-kindergarten reading readiness program aids early learning. She identifies 200 families that are eligible for the program, offers 100 the chance to participate and uses the other 100 families as a control group. So far so good. But suppose only 65 of the families invited to participate do so, and it later appears the children from these families are much better readers in first and second grade than the children in the control group. This difference may have emerged because the reading preparedness program really helped prepare the participating preschoolers to read, but it is also possible that the families that chose to participate were those most interested in having children who read well. The parents in these families may have been the kind who limited TV. watching and read to their children nightly. Such parents would have had children who were more likely than most to excel in early reading whether or not they had experienced the program. Had the

reading scores from the children of the thirty-five non-participating invitee families been counted along with the scores of the participants, the treatment and control groups might have been reading, on average, at the same level. Indeed, data from all families might suggest that the special program retarded reading.

For these reasons a common conservative practice in comparing treated and control group outcomes is to include post-test measures for all those randomly assigned to the treatment group regardless of whether they received the treatment. If the treatment had an effect, with enough treated cases it should still emerge from the data, although it would be weaker than it would have been with full participation. Alternatively, with adequate data there are ways of modeling propensities to participate that can help correct for participation bias.

To put this problem in the context of the proposed research, consider a situation where there are three students M(atched)1 and M(atched)2 (who are at a lower status law school and have identical admissions index credentials that are well matched to those of other students at that law school) and M(is)M(atched)3 who shares the same credentials but is at a higher ranking school and so is substantially mismatched in terms of entrance credentials. In determining whether mismatch has affected bar scores, we would want to compare the bar exam results of MM3 with the average of the bar exam scores of M1 and M2. Alternatively, if we were using an individual matching design we would randomly match MM3 with either M1 or M2. So long as our matched sample was large enough, we would expect that some of our random matches would be with the more able of the matching students and some with the less able, so on average differences caused by chance assignment of matches would cancel out.

Again, so far so good. But suppose M1 had failed in his first or second year, or left school because he didn't like legal analysis or felt that his grades, while passing, were not strong enough to justify the cost of completing his legal education. He may even have graduated but knowing the California bar was one of the nation's hardest decided to move to a state where he was more likely to pass the bar. M2, on the other hand, might have done well on law school exams and have confidence in her ability to pass the California bar. So she graduates and takes the California bar. Comparing her performance with that of MM3 in a mismatch study is, in layman's terms, *not fair*. More technically, it introduces a source of *bias* that might mistakenly suggest that a mismatch effect exists or even disguise a reverse mismatch effect. MM3's bar exam score might be significantly lower than that of M2, who can be matched only because she succeeded in her law school exams, success that predicts to doing well on the bar, but it might be significantly higher than the average score that would have been attained by M1 and M2 together had MI stayed in school, graduated and taken the bar.

This problem might not be as serious if drop out rates were consistent across California's law schools, but they are not. As school selectivity diminishes, drop out rates increase. This biases the study, especially the matching credential study, in favor of finding spurious mismatch effects. It can also hide reverse mismatch effects.

In a follow-up letter to the Bar, Sander attempted to respond to this point made by Kidder. The response is troublesome, to say the least. Rather than explore whether statistical corrections are possible, Sander attempts to minimize the problem. He does this first by separating documented 1L academic attrition from other attrition and assuming that other attrition either has no relationship to academic performance or a positive relationship. Thus Sander characterizes non-academic attrition as "overwhelmingly students who after a successful 1L year transfer to a more prestigious institution." The characterization is documented only by a vague reference to "the legal press."

Sander's attempt to minimize the statistical problems posed by the correlation of attrition with school status not only ignores attrition after the first year, but it also flies in the face of the BPS data that Sander is not reluctant to use to support the team's position. These data indicate that the most common reasons students give for failing to complete law school are financial, and they suggest that transfers are rare. Only 346 students of 27,320 students in the original BPS sample, or 1.3%, are reported as receiving their degree from a law school in a higher tier than the one they started out in. An additional 71 students transferred to a lower tier school. While there were presumably some transfers within tiers among the BPS respondents, unless California's law students are hugely different from the rest of the nation's, transfers in general and transfers to better schools in particular do little to minimize the threat that drop out rates that differ by school status pose for the research Sander proposes.

Sander's other response to this problem is that the BPS data reveal that students with similar credentials in different tier law schools graduate at similar rates. This claim undercuts the core claims in Professor Sander's original Stanford Law Review article: namely; that the BPS data show strong negative effects to mismatch, and without affirmative action enhanced black graduation and bar passage rates would produce more black lawyers than are produced through current admissions practices. I am pleased that Professor Sander seems now willing to repudiate at least part of the basis for his earlier arguments, but I still think Sander's effort to minimize selection bias problems is misleading.

In particular, because of what it says about how Sander is willing to describe data to defend a position, I was especially disturbed when I read, "[I]n some comparison groups, graduation rates for minorities with relatively high credentials are actually *higher* at less elite institutions."

Here is what the data in fact look like:

BLACK GRADUATION BY TIER AND INDEX DECILE							
DECILE	**ELITE**	**1ST TIER**	**2ND TIER**	**2ND TIER**	**Historic BLACK**	**3RD TIER**	**TOTAL**
1ST	**94.7% (18|1)**	**81.8% (81|18)**	**74.3% (296|92)**	**74.2% (216|75)**	**77.7% (202|58)**	**63.7% (58|33)**	
2ND	**100% (17)**	**83.9% (46|9)**	**85.3% (99|17)**	**82.8% (77|16)**	**81.3% (26|6)**	**83.3% (5|1)**	
3RD	**95.7% (22|1)**	**97.1% (33|1)**	**83.3% (50|10)**	**90% (27|3)**	**90.9% (10|1)**	*	
4TH	**91.7 (22|2)**	**90.5% (19|2)**	**92.3% (24|2)**	**90% (9|1)**	**90.9% (10|1)**	*	
5TH	**93.3% (14|1)**	**95% (19|1)**	**100% (14)**	**90% (9|1)**	*	*	
6TH	**100% (12|)**	**88.9 (16|2)**	**100% (12)**	**80% (4|1)**	*	*	
7TH	**100% (9)**	**100% (11)**	**100% (6)**	*	*	–	
8TH	**92.9% (13|1)**	**80% (4|1)**	*	*	–	–	
9TH	**100% (8)**	**100% (5)**	*	*	–	–	
10TH	*	*	–	–	–	–	

* Fewer than 5 students in cell, 3 of 4 elite students graduated, as did 3 of 3 prestige students, 4 of 4 2nd tier public students, 5 of 5 2nd tier private students, 6 of 7 historically black students, and 7 of 8 3rd tier students. In another table I and my coauthors generated in response to Professor Sander's earlier work, we broke the data down into demideciles allowing yet finer comparisons, but we were looking at the likelihood of graduation followed by bar passage rather than at just graduation. Patterns in that table are consistent with the results in the table above. Numbers of passers|fails in parentheses.

What one sees from this table is that black students in the BPS are overwhelmingly concentrated in the bottom two deciles of the index credential distribution. Indeed, only at the elite schools do we find more black students with index credentials above rather than below the second decile for all law students. Ignoring the historically black law schools, of which there are none in California, we see that in the bottom decile of the index distribution, where the majority of black students are found, elite schools have higher graduation rates than tier one schools, tier one schools have higher graduation rates than tier two schools and tier two schools have higher graduation rates than tier three schools. Among students with index credentials in the second decile elite schools maintain a graduation rate advantage over students in schools in other tiers, but the rate difference between these other schools vanishes. Above the second decile meaningful differences in graduation rates, except for a few perhaps anomalous results, largely disappear, but fewer than a quarter of the BPS black students had index scores in the upper 80% of the index score distribution. While those identified as mismatched in the California bar data will not be confined to black students, to the extent the BPS data can be relied on, as Sander did in the quoted sentence, they highlight rather than alleviate the danger of bias. As the above table indicates, differential drop out rates are likely to confound any attempt to study the mismatch hypothesis when data are available only for those who took the bar. *The need to graduate to enter the sample is very likely to bias results in the direction of finding mismatch*.

We also see from the table that Sander's statement "[I]n some comparison groups, graduation rates for minorities with relatively high credentials are actually *higher* at less elite institutions" is literally true. But this response to Kidder's methodological critique is troublesome, for it suggests the data look quite different than they do. It seems to me to be the argument of a lawyer who wields statistics in an effort to aid a client's cause and not the argument of a social scientist who uses statistical information to enlighten. With the possible exception of first decile students in historically black law schools, which are for black students atmospherically distinct from all other law schools, the index score controlled comparisons that show black students in lower tier schools as graduating at higher rates than black students in higher tier schools are based on differences so small in magnitude and/or on so few cases that no statement that attempts to minimize the potential selection bias problem by asserting that relationships are sometimes reversed is statistically justified.

A different point Kidder makes is that the number of minority graduates available in the sample years is so small that a meaningful before and after analysis of the effects of Proposition 209 would be difficult, if not impossible, to do. Difficulties are exacerbated, Kidder argued, because Hastings admissions policies before and after Prop. 209 differed so much from those of the other California public law schools, that their data should not be combined for analysis. Sander, responding to Kidder's memorandum, disputed Kidder's numbers and specifically rejected Kidder's argument about the need to treat Hastings separately. Not knowing the relevant data, I cannot weigh in on whose numbers are better, but I can say that Sander's argument for treating Hastings as if it had been affected no differently than California's other public law schools by Proposition 209

is not only unpersuasive but also disquieting in its attempt to wish away an issue that needs addressing.

According to the data Sander presented in his reply to Kidder, excluding Hastings the number of Hispanics, African Americans and American Indians matriculating at California's public law schools went from 482 in the three years before Prop 209 to 222 in the three years immediately following the measures adoption. To put this another way, the post-Prop 209 non-Asian minority enrollment in these schools is about 46% of what it was during the pre-Prop 209 base period. At Hastings during the same periods, Hispanic, African American and American Indian matriculation fell, according to Sander, from 166 to 121, leaving an enrollment that is about 73% of the original base. The difference between enrollment drop-off at Hastings and the other California public law schools is substantial and almost certainly statistically significant. This means that Prop 209 affected Hastings differently than it affected the other California public schools. The most likely reason is that Hastings minority admissions procedures were different from those of California's other public law schools either before Prop 209 or afterwards or both. This substantiates Kidder's caution against combining the Hastings data with the data from California's other public law schools, as Sander proposes to do.

The difference we see need not reflect a different approach to minority admissions if the Hastings falloff is relatively small because its pool of minorities who could be admitted without preferences grew after Prop 209. This is plausible because, post-Prop 209, Hastings could be attracting students within its ordinary admissions range who previously would have attended Boalt Hall or UCLA. But upon reflection, the possibility seems unlikely to account for most of the difference. Students who could have been admitted to Berkeley or UCLA before Prop 209 could still attend schools like Michigan, NYU, Cornell and Northwestern that are considerably more prestigious than Hastings. Moreover, if this cascade explanation were correct, it should also be true of Davis, but my understanding is that the drop in minority admissions that Davis experienced is closer to the experience of Boalt Hall and UCLA than it is to the experience at Hastings. If correct, this understanding together with Kidder's anecdotal evidence that Hastings differed from California's other public law schools in its approach to minority admissions, argues for treating Hastings graduates separately, as Kidder suggested, even at the cost of reduced sample size.

Remaining methodological issue raise what are probably smaller concerns. For example, what degree of mismatch will count as substantial mismatch, what differences in bar test results will be regarded as important, etc? Most data sets contain substantial random error. If decisions on issues like those raised by the above questions are based not on theory or otherwise specified *a priori* but are determined only after analyzing the data, the validity of apparently significant differences may be questionable and other dangers arise. Hence, an additional reason to require a more detailed proposal from Sander if the Bar can legally provide Sander the data in its possession is so that the Bar can know Sander's thoughts on issues like these before it reaches an access decision. In this connection the Bar should inquire whether Dr. Klein has already looked informally or in other research at some of the relationships that would be the object of study.

An additional concern involves possible gaps in the data the Bar's possesses. It appears that some of the data Sander would use is provided by bar applicants. This suggests a need to evaluate reliability and missing data problems. The Bar needs to know the extent of missing data or other data reliability problems in order to evaluate possible payoffs from the proposed research. I assume Dr. Klein has this information.

A final research-related issue, though not a methodological one, involves the funding for this research. The original proposal contemplated NSF funding but indicated that other sources of funding were being explored. With NSF funding out of the picture, the Bar should inquire into how the proposed research will be funded and what the contemplated budget will be. While funding sources have no necessary implications for the quality or unbiasedness of work done, NSF funding would have added credibility to the research effort while one can imagine funding sources that would have the opposite impact.

Data Access

The Bar should regard a decision to provide Sander access to the data he requests as a commitment to provide similar access to other legitimate researchers. To my knowledge no published paper looking at the data Professor Sander used in his Stanford Law Review articles confirmed his claim to have found a substantial mismatch effect. Moreover, even in this commentary we have seen several occasions where an independent look at data raised questions about assertions Sander made.

Sander does not dispute the desirability of allowing other researchers to have access to the data. Rather, if I read what the Sander team has said correctly, I think, to its credit, it endorses this. But none of the suggestions that Sander makes for giving independent researchers broader access to the Bar data is adequate. Forcing other researchers to work through Drs. Klein and Bolus is unworkable if, for no other reason, than, as Sander notes in his memorandum of November 19, 2006, researchers using Klein and Bolus's services would have to have funds to pay for them. Not only is money hard to come by for replication research, but even if funds were acquired Klein and Bolus might not have the time or interest to participate in further analysis. Moreover, much of the work that one does to see whether another's results and interpretations stand up to scrutiny is exploratory in nature. Working through a third party would allow specific tests to be run, but would not allow the kind of broad exploratory analysis that proved necessary to evaluate Prof. Sander's past work.

I might add that despite the statistical competence of Dr. Klein and Dr. Bolus relying on them for all analyses done by the Sander team is a further weakness of the proposal. If this project is to proceed, it would be strengthened if all research team members could examine the data. That way it could benefit from the internal methodological critique and insights that someone like Bill Henderson, or the economist on the team whom I do not know, might provide. Moreover including on the project

team people like Amar and Henderson, who do not have Sander's reputation (fair or not) for opposing affirmative action, does not add much to the credibility of the research if they are removed from the data analysis.

Using scholars associated with JELS as gatekeepers is a somewhat odd suggestion, especially as one member of the Sander team, Professor Henderson, has had, unless I am mistaken, a close relationship to JELS. Nevertheless I regard Professor Henderson as a scholar of considerable integrity and regard others who are associated with JELS, like Professor Eisenberg, in the same way. I am confident that Professor Henderson would distance himself from decisions regarding further use of the data and indeed would trust him to participate in such decisions even if it was his own work that would be examined. So it is not a concern for fairness that leads me to question the JELS option. Rather it is the idea that third parties should evaluate the quality of proposed research and allow a study only if it passes some test.

Evaluating a proposal's quality and likely knowledge payoff is appropriate for a funder or for an organization, like the Bar, that must decide whether to allow its data to be used in research, and it is, of course, appropriate for a journal when it is making a publication decision. The situation is different, however, when the data will be used to examine the research findings of others. It may be impossible to specify in advance where the investigation may lead or even what methods will be used. Decisions on such matters may depend on what exploratory analysis of the data reveals. If a third party data gatekeeper is to play a role, it should be limited to ensuring that those seeking the data are bona fide social scientists who have received appropriate human subjects clearance from their home institutions. Also giving the data to one scholar should not preclude another from assessing it as well. Once the Bar offers its data to one scholar, it should remain concerned that the anonymity, privacy and related interests of its test takers will be guaranteed and that its data will be used for legitimate research purposes. But it should also be concerned that the original research its data enabled stands up to scientific scrutiny. There is no research purpose more legitimate than testing the claimed findings of others, and few purposes are more important when purported findings bear on sensitive, politically charged issues.

Conclusion

. It may be that the data in the Bar's possession cannot lawfully be used for the project proposed by the Sander team due to the lack of informed subject consent or dangers posed to human subjects. It may also be that the Bar's internal rules or state or federal law preclude the use of the bar test data for purposes like those proposed. The first issue seems to me to be for an IRB to first decide, though the Bar may have its own special concerns which cannot be delegated to the discretion of others. The second raises issues for the Bar Board of Governors or its legal counsel.

As a researcher and law professor, I hope that barriers to using bar exam data for purposes beyond narrowly bar-relevant studies are not insurmountable, for I can imagine

studies using bar test data that might, if linked with data from law schools or bar prep courses, productively inform legal education. However, as a lawyer I also recognize the legitimacy of rules designed to protect the interests of the bar and its applicants. Hence I think there is a heavy burden to be borne by those who suggest the Bar Examiners' decision was made in bad faith. If, however, the Bar Governance Board feels its data *can* be used as Sander proposes, I suggest the Bar approach the decision of whether it *should* release its data by addressing, as rigorously as possible, two concerns.

The first is whether publications from the proposed research, even if they do not reveal applicant identities, will nonetheless adversely affect applicant privacy or wrongly imply that certain groups of bar passers have learned more in law school and/or are likely to be more capable attorneys than others. The second is whether the quality of the available data and the measures they allow are adequate to the research task. To make the latter assessment the Bar must determine whether the proposal adequately anticipates and confronts some of the difficult methodological hurdles that design and data quality issues suggest. To do this, the Bar, in my view, would be well advised to ask a small group of top social science methodologists to evaluate the Sander proposal. I also suggest that if the Bar takes this path, Sander would be well advised to rewrite the team's proposal rather than simply supplement it with additional memoranda. A fresh proposal should candidly confront the issues Kidder, I, some of the NSF reviewers and others have identified in order to make the proposal as scientifically strong as possible. Where arguments from existing data are made, Sander should provide considerable detail about what relevant data show and not merely advance the data or interpretations that seem to best support the team's positions.

My hunch is that the hurdles that data quality, measurement, privacy, informed consent and design issues place on usefully informative research are too great to surmount, but I hope that I am wrong. Evaluating the mismatch hypothesis is a legitimate social science concern and has obvious relevance to important policy issues. Indeed, if the data set is of sufficient quality, and if design problems like selection bias and obstacles like the lack of informed consent can be overcome, I myself would like to have access to the data used by Sander's team.

I also hope I am wrong in what is a fear more than a hunch; namely that the Bar Board of Governors will decide whether to approve the use of its data not on the basis of its evaluation of the privacy, informed consent and quality issues that should determine its answer to Sander but rather because of pressures placed on the Board to let this research proceed. Something is wrong when members of the politically powerful United States Commission on Civil Rights weigh in on a local issue like this, and things seem even more wrong when the only Commission members who write to urge the board to overturn the Bar Examiners' decision are known opponents of affirmative action. The Bar, I am sure, is acutely aware that some think the proposed research is opposed only because its results may not be "politically correct," while others fear that data access may be given because the only politically correct response to being called politically correct is to do what the name callers say you won't.

My view is that the Board of Governors should put these thoughts aside. Its decision should be influenced only by legal, ethical and scientific considerations and not turn on its members' political preferences or a felt need to bow to political pressures, whatever direction they come from. Of these three concerns only questions relating to the law and bar rules are ripe for decision now. Ethical issues should be first addressed by the UCLA IRB, and only a substantially revised proposal might alleviate the scientific concerns described above.

I apologize for the length of the memorandum. In its defense, I can say that its length is a signal that the issues to be worked out before this proposal can proceed are many, complex, and require thoughtful consideration by trained ethicists and social scientists. I appreciate your patience. If you have any questions regarding the content of this memorandum or the Sander team's proposal generally, please do not hesitate to contact me at rlempert@umich.edu. I will provide a phone number upon request.