October 1, 2017

The Honorable Tani G. Cantil-Sakauye
   and Associate Justices
Supreme Court of California
350 McAllister Street
San Francisco, California 94102

Docket Number S244281: In re California Bar Exam

Dear Chief Justice Cantil-Sakauye and Associate Justices:

Thank you for this opportunity to comment on the Court's consideration of an appropriate cut score for California's bar exam. I am a legal educator who has studied and written about psychometric properties of the bar exam. I also work frequently with statistical techniques in my research and have taught courses explaining those techniques to law students. Because of this work, I believe that I can offer useful insights into the social science basis of the proposals contained in the State Bar of California's *Final Report on the 2017 California Bar Exam Standard Setting Study* and in the underlying *Standard Setting Study* prepared by ACS Ventures.

From a social science perspective, there are three key facts about California's cut score and the recent standard-setting study:

First, California's bar exam—like the exams in other states—has never been validated. This means that no job analysis or other scientific study links the exam's content to the skills and knowledge needed by new attorneys. At the same time, California has set an unusually high cut score with a disproportionate racial impact: white test-takers are much more likely to pass California's exam than test-takers of any other race or ethnicity. The combination of a non-validated exam, an unusually high cut score, and a racially disproportionate impact is very troubling from both a social science and policy perspective.

Some participants in the cut-score debate have suggested that the Court should preserve the current cut score until the exam has been validated through a thorough job analysis. Given California's diversity and the score's racially disproportionate impact, however, just the opposite is true. California should not maintain a high cut score that excludes a disproportionate number of its minority residents from practicing law unless it can *first* validate the exam and link the high cut score to minimal competence in practicing law. A racially disproportionate impact should be justified; it should not be the default.

Second, although California's recent standard-setting study was well intentioned, it does not provide a scientifically valid foundation for the cut scores recommended by ACS Ventures or the

State Bar. As I explained in a memo submitted to the State Bar (attached here for reference), the scores underlying the ACS calculations lack coherence; the participants in the study identified essays with widely divergent scores as minimally competent. These scattered results do not support a valid cut score.

In its *Final Report* to this Court, the State Bar acknowledges this problem of scattered scores. The *Report* concedes that Dr. Nyblade (a distinguished empiricist who commented on this same problem) and I were "correct in terms of the dispersion of the data that they reviewed." *Final Report* at 30. The *Report* then attempts to avoid this problem by asserting that a different set of scores—the raw scores produced the panelists—were more concentrated. That assertion, however, is incorrect.[1]

As the "leaf" diagram on p. 32 of the State Bar's *Final Report* shows, panelists chose papers of widely varying quality as borderline passing papers. It is easier to see the range of disagreement in a simple bar graph than in the more unusual leaf format:
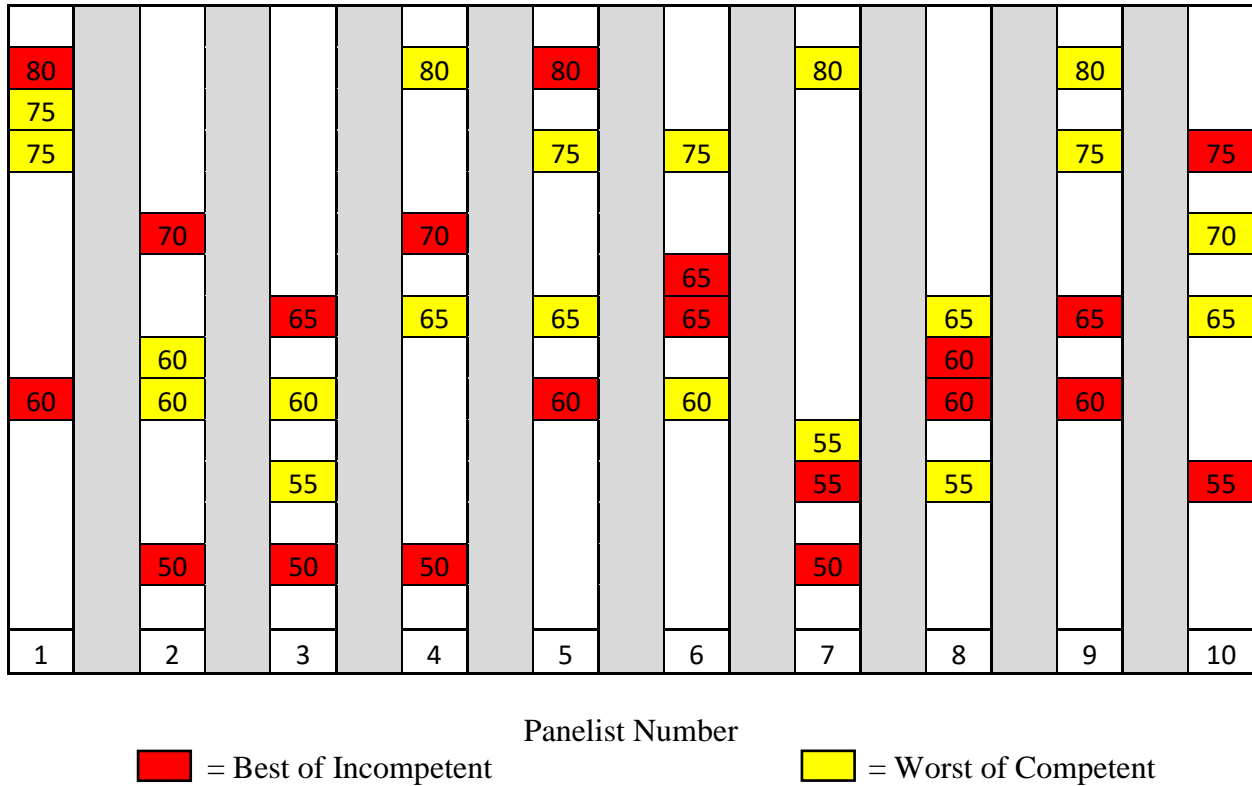

Graph 1: Scores of Borderline Essays
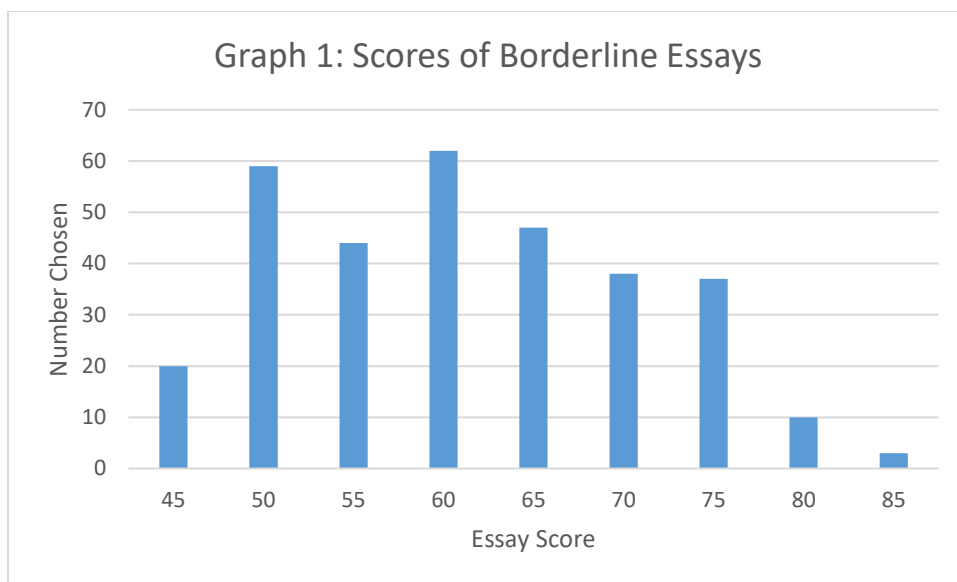
As this graph shows, the panelists chose essays from every score category as the "best of the incompetent" or "worst of the competent" essays they reviewed. On thirteen occasions, in fact, a panelist chose a very high-scoring essay (one awarded a grade of 80 or 85) as a borderline essay. This is particularly remarkable because ACS Ventures included only a small number of those high-scoring essays in the standard-setting exercise.

---

[1] The assertion is also problematic because ACS Ventures did not use these raw scores to calculate the cut scores proposed in its report. Using the raw scores identified by the State Bar in its *Final Report* would generate a *lower* set of cut scores than the ones currently endorsed by ACS Ventures and the State Bar. The median value for the set of 320 raw scores discussed by the State Bar in its *Final Report* is 60. That value produces a cut score of 420 (rather than 425) for the written part of the exam. These inconsistencies reduce confidence in the recommendations embodied in either report.

Equally troubling, many of the panelists made choices that were internally inconsistent. As the graph below illustrates, panelists frequently chose a high-scoring essay as an incompetent paper, while marking a lower-scoring essay as a competent one. In the graph, each column represents the four borderline essays chosen by each panelist for one of the questions assessed during the exercise. Yellow blocks represent essays that the panelist thought were competent; red blocks represent essays that the panelist thought were incompetent.

Graph 2: Borderline Essays Identified for Question Two

Score	1	2	3	4	5	6	7	8	9	10
80	red			yellow	red		yellow		yellow	
75	yellow				yellow	yellow			yellow	red
75	yellow									
70		red		red						yellow
65			red	yellow	yellow	red		yellow	red	yellow
65						red				
60	red	yellow	yellow		red	yellow		red	red	
60		yellow						red		
55			yellow				yellow / red	yellow		red
50		red		red			red			

Panelist Number

■ = Best of Incompetent   ■ = Worst of Competent

If the panelists were consistent in their scoring, the red blocks (incompetent essays) would appear below the yellow (competent) ones. Instead, the scores produce a disorganized quilt-like pattern. Four panelists identified essays that had been awarded 50 points as incompetent; two put essays that had received 80 points in that category. Eight of the ten panelists were inconsistent in their own ratings: they rated at least one low-scoring essay as more competent than a higher-scoring one. Similar variations occurred in the assessment of other questions.

Panelists in any standard-setting exercise will disagree over what score signals minimum competence: some panelists will favor a higher score than others. It is unusual, however, for panelists to show this range of disagreement—or for their individual scores to be so internally inconsistent. The widely varied scores suggest that the group lacked a coherent vision of minimal competence.

This lack of coherence raises two troubling possibilities. On the one hand, the legal profession may lack a coherent view of the minimal competence measured on the bar exam. If that is true, then the exam itself needs restructuring. Alternatively, the panelists may have shared a vision of minimal competence—but this study was too flawed to tap that concept accurately. In either case, the scores generated by the study do not provide a valid basis for setting a cut score.[2]

Third, ACS Ventures and the State Bar improperly translated a cut score based on just the written portion of the exam to one that would encompass both written and multiple-choice portions of the test. ACS Ventures accomplished that goal by using a statistical technique known as equipercentile linking. As the ACS report recognizes, however, this technique assumes that scores on the written and multiple-choice portions of the exam "are sufficiently correlated to support" the translation. *Standard Setting Study*, at 14.

ACS Ventures then cites an outdated correlation that, I believe, misled the State Bar. ACS notes that "the correlation between the written scores and the total [exam] score . . . was 0.97" on the July 2016 bar exam. *Id.* The July 2016 exam, however, weighted written scores as 65% of the final exam score; under those circumstances, a high correlation between written scores and final score is expected. California, however, now weights written and multiple-choice scores equally. No one will know the correlation between written and total scores on the current exam until California finishes grading that exam, but the correlation undoubtedly will be lower than 0.97.

At the same time, ACS Ventures overlooked a much more relevant correlation. The *Bolus Report*, prepared earlier this year for the State Bar, finds that the correlation between written and multiple-choice scores on the California bar exam ranges from 0.55 to 0.73.[3] These correlations are not sufficiently strong to support the equipercentile linking approach that ACS Ventures used to calculate a proposed cut score for the full exam.

Neglecting to account for these lower correlations is especially troublesome because California, like most states, grades the bar exam on a compensatory basis. Test-takers do not have to establish minimal competence on both the multiple-choice and written portions of the exam; they must establish minimal competence through their combined score. A high score on the multiple-choice questions, in other words, may compensate for a low score on the written essays (and vice versa). The statistical technique used by ACS Ventures, and adopted by the State Bar, overlooks both the relevant correlations and the compensatory nature of California's grading.

Regrettably, the State Bar does not mention this problem in its *Final Report*, although I (and perhaps others) raised the issue in comments submitted to the Bar. The use of the misleading 0.97

---

[2] ACS Ventures and the State Bar suggest that their recommended cut scores are *reliable*, because a statistical test suggests that repeating the standard-setting study (in its current form) would generate similar means and medians. Reliability, however, is quite different from validity. Reliability reflects the ability to replicate a particular result; it does not indicate that the result is a good measure of the underlying concept. If scientists repeated the study conducted by ACS Ventures, including the flaws embodied in that study, they almost certainly would obtain widely scattered scores like the ones produced in the current study. One could, once again, calculate a median from those scattered scores—and the median might be similar to the one obtained here. Calculating that median, however, would be just as invalid the second time as the first. One must establish a study's validity as well as its replicability.

[3] ROGER BOLUS, RECENT PERFORMANCE CHANGES ON THE CALIFORNIA BAR EXAMINATION: INSIGHTS FROM CBE ELECTRONIC DATABASES 26 (2017).

correlation and the neglect of California's compensatory scoring are not minor technical issues. These flaws undermine any scientific basis for the cut scores recommended by the State Bar. At the very least, they counsel that the Court should adopt a score *below* the ones recommended by the State Bar. That approach would help account for the compensatory grading that ACS Ventures and the State Bar overlooked.

As a social science scholar and teacher, I embrace policy decisions informed by social science. The science, however, has to be well grounded before policy makers draw upon it. In this case, despite a good faith attempt to generate a scientifically based cut score, the work by ACS Ventures fell short. Valid social science studies are difficult to conduct; a good study takes many months (or years) to design and implement. I hope that California will continue its leadership in studying the bar exam scientifically, but the current study should play a minimal role in the Court's decision making.

Given the serious flaws in the standard-setting study, the first principle discussed above assumes particular importance: A state should not maintain an unusually high cut score with a racially disproportionate impact unless it can point to rigorous studies validating both the exam's content and the unusual cut score. Absent those studies, the best course is to adopt an interim cut score similar to the ones embraced by a majority of other states. Those cut scores have produced no negative effect on public welfare and they are sufficient to maintain public confidence in the bar. Meanwhile, I hope that California and other states will continue studying the validity of the bar exam.[4]

I hope that my comments are helpful to the Court. Very few of my students take the California Bar Exam, but I have a deep interest in the psychometrics of professional licensing exams and the broader application of social science techniques to policy issues. Please let me know if I can provide any further assistance.

Sincerely yours,

Deborah Jones Merritt
John Deaver Drinko
Baker & Hostetler Chair in Law
merritt.52@osu.edu

---

[4] *See* Deborah Jones Merritt, *Validity, Competence, and the Bar Exam*, AALSNEWS, Spring 2017, at 11, *available at* https://www.aals.org/wp-content/uploads/2017/05/AALSnews_spring17-v9.pdf.