

EXHIBIT A

Proof 100% Subjective
Cal Bar Exam for
experienced attorneys is
not a *valid* or *reliable* test
covered-up by Federal
Courts



**Committee of Bar Examiners
of The State Bar of California**

**Report to the
Supreme Court of California
on the February 2001
California Bar Examination**

WRITTEN SECTION

There were 1050 applicants who had their answers read at least twice. On the average, an applicant's total written raw score on the first reading was about 5 points higher than it was on the second reading. The correlation between these scores was .41. This value underestimates the true degree of agreement between readers because reread was limited to applicants near the pass/fail line. Table 2 shows the means and standard deviations on each question after all readings.

Table 2 - SUMMARY STATISTICAL DATA ON THE WRITTEN SECTION AFTER ALL READINGS

Question Number	Essay Content Area(s) and PT Tasks	Mean Score	Standard Deviation
1	Community Property	61.7	8.32
2	Corporations	63.3	7.32
3	Criminal Law	59.8	8.50
4	Remedies	64.3	7.03
5	Professional Responsibility	65.6	7.12
6	Contracts	61.3	6.39
PT-1	Draft pre-counseling letter	65.3	8.57
PT-2	Draft closing argument	66.8	7.86

SUMMARY STATISTICS

Table 3 presents summary statistical data on each section after all readings. There was a .58 correlation between MBE and Written scores. Law School Admission Test (LSAT) scores correlated .52, .43, and .52 with MBE, Written, and Total Scale scores, respectively. There were 3779 applicants with useable LSAT scores.

Table 3 - SUMMARY TEST STATISTICS AFTER ALL READINGS

Test Statistic	MBE Scale	Written Raw	Total Scale
Mean Score	1405.3	640.2	1397.5
Standard Deviation	147.0	46.2	129.5
Reliability	.89	.72	.83

The MBE's reliability was computed by ACT using national data.



**THE COMMITTEE OF BAR EXAMINERS
OF THE STATE BAR OF CALIFORNIA**

**Report to the
Supreme Court of California
on the
July 2001 California Bar Examination**

(July 2001 .48)Table 1 - NATIONAL AND CALIFORNIA MEAN MBE SCORES
AND THE DIFFERENCE BETWEEN THESE MEANS

Test Score	Number of Items	National Mean	California Mean	Difference
Constitutional Law	33	22.9	22.5	-0.4
Contracts	34	21.8	22.3	0.5
Criminal Law	33	21.6	22.5	0.9
Evidence	33	21.4	23.5	2.1
Real Property	33	19.2	20.5	1.3
Torts	34	21.9	22.4	0.5
Total Raw	200	128.7	133.7	5.0
NCBE/ACT Scale	200	142.8	146.8	4.0

WRITTEN SECTION

There were 1598 applicants who had their answers read at least twice. On the average, an applicant's total written raw score on the first reading was 7 points higher than it was on the second reading. The correlation between these scores was .48. This value underestimates the true degree of agreement between readers because reread was limited to applicants near the pass/fail line. Table 2 shows the means and standard deviations on each question after all readings.

Table 2 - SUMMARY STATISTICAL DATA ON THE
WRITTEN SECTION AFTER ALL READINGS

Question Number	Essay Content Area(s) and PT Tasks	Mean Score	Standard Deviation
1	Civil Procedure	61.7	9.19
2	Real Property	64.9	7.22
3	Evidence	64.1	9.45
4	Constitutional Law	61.4	7.92
5	Torts	61.3	8.91
6	Wills & Trusts	61.8	7.57
PT-1	Constitutional Law	66.1	6.61
PT-2	Criminal Law and Procedure	65.0	8.11



**THE COMMITTEE OF BAR EXAMINERS
OF THE STATE BAR OF CALIFORNIA**

**Report to the
Supreme Court of California
on the
February 2002
California Bar Examination**

(February 2002 .38)

Table 1 - NATIONAL AND CALIFORNIA MEAN MBE SCORES
AND THE DIFFERENCE BETWEEN THESE MEANS

Test Score	Number of Items	National Mean	California Mean	Difference
Constitutional Law	33	20.71	18.03	- 2.68
Contracts	34	22.83	18.72	- 4.11
Criminal Law	33	17.08	23.84	6.76
Evidence	33	17.94	21.12	3.18
Real Property	33	19.65	20.43	0.78
Torts	34	21.93	22.95	1.02
Total Raw	200	120.15	125.09	4.94
NCBE/ACT Scale	200	135.28	139.60	4.32

WRITTEN SECTION

There were 990 applicants who had their answers read at least twice. On the average, an applicant's total written raw score on the first reading was 11 points higher than it was on the second reading. The correlation between these scores was .38. This value underestimates the true degree of agreement between readers because reread was limited to applicants near the pass/fail line. Table 2 shows the means and standard deviations on each question after all readings.



THE COMMITTEE OF BAR EXAMINERS
OF THE STATE BAR OF CALIFORNIA

**Report to the
Supreme Court of California
on the
July 2002 California Bar Examination**

(July 2002 .40)**WRITTEN SECTION**

There were 1588 applicants who had their answers read at least twice. On the average, an applicant's total written raw score on the first reading was 4 points higher than it was on the second reading. The correlation between these scores was .40. This value underestimates the true degree of agreement between readers because reread was limited to applicants near the pass/fail line. Table 2 shows the means and standard deviations on each question after all readings.

Table 2 - SUMMARY STATISTICAL DATA ON THE WRITTEN SECTION AFTER ALL READINGS

Question Number	Essay Content Area(s) and PT Tasks	Mean Score	Standard Deviation
1	Wills	63.3	10.2
2	Real Property	63.5	6.8
3	Professional Responsibility	65.3	9.2
4	Contracts	62.4	8.2
5	Torts	61.2	8.3
6	Community Property	63.2	7.7
PT-1	Memo regarding constitutionality and changes	60.8	8.8
PT-2	Analysis of criminal law statute and ethical issues	65.8	7.4

SUMMARY STATISTICS

Table 3 presents summary statistical data on each section after all readings. There was a .64 correlation between MBE and Written scores. Law School Admission Test (LSAT) scores correlated .61, .48, and .57 with MBE, Written, and Total Scale scores, respectively. There were 6764 applicants with useable LSAT scores.

Table 3 - SUMMARY TEST STATISTICS AFTER ALL READINGS

Test Statistic	MBE Scale	Written Raw	Total Scale
Mean Score	1445.1	1438.3	1440.7
Standard Deviation	155.2	154.2	141.2
Reliability	0.89	0.72	0.84

The MBE's reliability was computed by ACT using national data.



**THE COMMITTEE OF BAR EXAMINERS
OF THE STATE BAR OF CALIFORNIA**

**Report to the
Supreme Court of California
on the
February 2003
California Bar Examination**

(February 2003 .48)**WRITTEN SECTION**

There were 991 applicants who had their answers read at least twice. On the average, an applicant's total written raw score on the first reading was 10 points higher than it was on the second reading. The correlation between these scores was .48. This value underestimates the true degree of agreement between readers because reread was limited to applicants near the pass/fail line. Table 2 shows the means and standard deviations on each question after all readings.

Table 2 - SUMMARY STATISTICAL DATA ON THE WRITTEN SECTION AFTER ALL READINGS

Question Number	Essay Content Area(s) and PT Tasks	Mean Score	Standard Deviation
1	Civil Procedure	55.56	8.11
2	Wills/Real Property	59.29	8.30
3	Criminal Law	61.10	6.93
4	Professional Responsibility	62.48	8.73
5	Constitutional Law	60.09	7.00
6	Community Property	59.20	7.97
PT-1	Torts	65.33	7.69
PT-2	Civil Procedure/Evidence	61.31	8.16

SUMMARY STATISTICS

Table 3 presents summary statistical data on each section after all readings. There was a .58 correlation between MBE and Written scores. Law School Admission Test (LSAT) scores correlated .48, .44, and .51 with MBE, Written, and Total Scale scores, respectively. There were 3523 applicants with useable LSAT scores.

Table 3 - SUMMARY TEST STATISTICS AFTER ALL READINGS

Test Statistic	MBE Scale	Written Raw	Total Scale
Mean Score	1397.92	611.06	1397.15
Standard Deviation	146.39	45.21	131.60
Reliability	.87	.68	.81

The MBE's reliability was computed by ACT using national data.



THE COMMITTEE OF BAR EXAMINERS
OF THE STATE BAR OF CALIFORNIA

**Report to the
Supreme Court of California
on the
February 2004
California Bar Examination**

WRITTEN SECTION

There were 936 applicants who had their answers read at least twice. On the average, an applicant's total written raw score on the first reading was 9 points higher than it was on the second reading. The correlation between these scores was .39. This value underestimates the true degree of agreement between readers because reread was limited to applicants near the pass/fail line. Table 2 shows the means and standard deviations on each question after all readings.

Table 2 - SUMMARY STATISTICAL DATA ON THE WRITTEN SECTION AFTER ALL READINGS

Question Number	Essay Content Area(s) and PT Tasks	Mean Score	Standard Deviation
1	Criminal Law and Procedures	62.90	8.05
2	Community Property	64.44	8.94
3	Professional Responsibility	64.38	7.64
4	Real Property	61.03	7.34
5	Constitutional Law	60.50	7.13
6	Civil Procedure	61.16	8.90
PT-1	Insurance Law and Contracts	61.70	7.80
PT-2	Contracts and Alternative Dispute Resolution	63.34	7.30

SUMMARY STATISTICS

Table 3 presents summary statistical data on each section after all readings. There was a .59 correlation between MBE and Written scores. Law School Admission Test (LSAT) scores correlated .52, .41, and .50 with MBE, Written, and Total Scale scores, respectively. There were 3699 applicants with useable LSAT scores

Table 3 - SUMMARY TEST STATISTICS AFTER ALL READINGS

Test Statistic	MBE Scale	Written Raw	Total Scale
Mean Score	1392.30	624.54	1390.27
Standard Deviation	146.53	46.28	132.04
Reliability	.89	.72	.90

The MBE's reliability was computed by ACT using national data.



THE COMMITTEE OF BAR EXAMINERS
OF THE STATE BAR OF CALIFORNIA

Report to the
Supreme Court of California
on the
July 2004
California Bar Examination

(July 2004 .41)**MULTISTATE BAR EXAMINATION (MBE)**

Table 1 shows California applicants scored higher than the national average on five of the six MBE subtests. California's mean total raw score (the average number of questions answered correctly) was 2.44 points higher than the national average (which included California scores).

Table 1 - NATIONAL AND CALIFORNIA MEAN MBE SCORES AND THE DIFFERENCE BETWEEN THESE MEANS

Test Score	Number of Items	National Mean	California Mean	Difference
Constitutional Law	33	23.51	21.63	-1.88
Contracts	34	21.63	21.97	0.34
Criminal Law	33	21.23	21.73	0.50
Evidence	33	21.65	23.96	2.31
Real Property	33	21.64	22.21	0.57
Torts	34	22.95	23.56	0.61
Total Raw	200	132.62	135.06	2.44
NCBE/ACT Scale	200	141.22	143.38	2.16

WRITTEN SECTION

There were 1658 applicants who had their answers read at least twice. On the average, an applicant's total written raw score on the first reading was 2.5 points higher than it was on the second reading. The correlation between these scores was .41. This value underestimates the true degree of agreement between readers because reread was limited to applicants near the pass/fail line. Table 2 shows the means and standard deviations on each question after all readings.

Table 2 - WRITTEN RAW SCORES AFTER ALL READINGS

Question Number	Essay Content Area(s) and PT Tasks	Mean Score	Standard Deviation
1	Criminal Law	61.08	6.79
2	Constitutional Law	59.27	6.90
3	Wills/Trusts	60.92	9.05
4	Evidence	62.85	8.65
5	Professional Responsibility	64.37	9.65
6	Torts	61.03	8.16
PT-1	Consumer Protection/Usury	60.86	7.72
PT-2	Torts/Premises Liability	64.84	8.35

DECLARATION OF PHILLIP L. ACKERMAN, Ph.D.

I PHILLIP L. ACKERMAN, Ph.D. am competent to testify and have personal knowledge, and based on that knowledge declare the following:

1. My curriculum vita is attached.

I am attaching a true and complete copy of my Evaluation of the Psychometric Adequacy of the California Attorney's Examination dated February 15, 2008, that is the test given by the State of California to attorneys in good standing already licensed in another state for four years to qualify for licensure in California. I am re-publishing that Evaluation, and including additional information that can be found in paragraphs 6 - 10 below, that reinforces the conclusions reached in my Evaluation of the Psychometric Adequacy of the California Attorney's Examination.

2. These are a few salient details about my background (also listed in the CV).

a. I am a Professor of Psychology at Georgia Institute of Technology, and the past Editor, *Journal of Experimental Psychology: Applied*.

b. I am a Fellow of the American Psychological Association, a Fellow of the American Educational Research Association, and a member of the National Council on Measurement in Education (these are the three organizations that generate the *Standards on Psychological and Educational Testing* in the United States.)

c. Over the past 17 years, I have published 12 reviews and I have two additional reviews "in press" in the *Mental Measurements Yearbook* (this is generally regarded as the 'bible' for critical reviews of commercial educational, psychological and organizational tests). In 2005, I was named a "Distinguished Reviewer" by the Buros Institute of Mental Measurements.

d. Over the past 27 years, I have consulted on educational and occupational testing for the following organizations: U.S. Air Force, U. S. Army, U. S. Navy Personnel Research and Development Center, U. S. Department of Education, Minnesota Air Traffic Control Training Center (FAA), The College Board, Educational Testing Service (ETS), and General Motors.

3. It is my professional opinion that the Attorney's Examination fails to meet the *Standards for Educational and Psychological Testing (1999)* published by the American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education.

4. As stated in my Evaluation of the Psychometric Adequacy of the California Attorney's Examination dated February 15, 2008, the specific shortcomings for this test and thus its failure to meet the *Standards for Educational and Psychological Testing* are as follows:

- a. The Attorney's Examination lacks content-related validity. According to Steven P. Klein (The Rand Corporation):
"State bar examinations have been criticized for measuring only a few of the important skills and abilities that are needed for the practice of law. For example, a typical essay question provides several facts that are material to a case and then asks the applicant to determine how the case should be resolved relative to the applicant's knowledge of general legal principles. **The exam does not address interviewing, negotiating, or oral advocacy skills; the ability to draft or evaluate legal documents; or the ability to conduct legal research.**" (Klein, 1983a; *Measuring Legal Research Skills on a Bar*

Examination, Rand Report P-6879). [emphasis added]¹

Similarly, Klein claimed “Some of the other skills that are important for legal practice that are not tested directly by the traditional bar exam include the ability to interview clients, examine a witness, conduct legal research, and negotiate a settlement.” (Klein, 1983b, *Relationship of Bar Examinations to Performance Tests of Lawyering Skills*, Rand Report P-6895)

- b. The Attorney’s Examination has never been demonstrated to have criterion-related validation, in terms of evaluating the scores on the test and comparing them to performance of practicing attorneys. (per Dr. Susan M. Case, Director of Testing, National Conference on Bar Examiners, 1/18/08).
- c. The scores on the Attorney’s Examination are determined in a manner that is not consistent with professional standards. The reliability of the test scoring procedures fails to reach a level that would be acceptable for high-stakes testing. (Specifically, inter-rater agreement is quite low, a correlation of .48 between raters indicate only 23% shared variance among ratings; source: Klein & Bolus; Gansk & Associates 2003.) An acceptable level of reliability for such high-stakes testing would be shared variance in the neighborhood of 70% or higher (corresponding to reliability of about .84 or higher).
- d. The passing cut-off score on the Attorney’s Examination is determined without regard to the actual knowledge and skills of the individuals taking the examination, but is determined by the scores of other individuals taking the test.

The adjustment process that is used to transform the Attorney’s Examination raw scores to a mean and standard deviation that matches the results of other examinee’s MBE scores, is performed in a manner that is inconsistent with actually assessing the individual examinee’s knowledge and skills. This process is “. . . arbitrary, because it ensures that some fixed proportion of applicants will fail even though all or most of the applicants may in fact be qualified. The more able the group of applicants taking the test, the higher the passing scores will be.” (Shimberg, 1981, p. 1141)

- e. Because of the lower reliability of the Attorney’s Examination, when compared to the MBE, and the non-compensatory scoring for the Attorney’s Examination, the resulting regression-to-the-mean (a statistical phenomena that is exacerbated by lower reliability) will result in fewer qualified individuals taking the Attorney’s Examination actually obtaining a passing score on the examination, in comparison to the California Bar Examination.
- 6. I have not testified as a witness at trial or deposition in the last four years.
- 7. I agree with Dr. Robert Kane’s conclusions that bar exams are high-stakes licensing tests, and “a fairly high reliability (above 0.8; preferably above 0.9) is expected for testing programs that are used to make high-stakes decisions about individuals.” See Kane,

¹ The incomplete coverage of the content of the job in this case might be compared to a driving license examination that only involves parking a car, and does not involve driving outside of a parking lot. Such an examination would fail to adequately sample the content of the overarching activities allowed by a driver’s license.

Reflections on Bar Examining, 6, The Bar Examiner, p.9 (Nov 2009). Dr. Kane's opinion is consistent with my opinion expressed in Paragraph 4C above: ("An acceptable level of reliability for such high-stakes testing would be shared variance in the neighborhood of 70% or higher (corresponding to reliability of about .84 or higher").

8. I also agree with Dr. Gary McClelland's conclusion concerning the California Attorney's Examination that: "The degree of inter-rater agreement is dreadful. I do not believe any scientist would ever publish data based on such low inter-rater agreement." See McClelland Declaration ¶ 6, May 5, 2008.

9. When the goal of inter-rater reliability is preferably in the range of .8 to .9 as noted by Dr. Kane, and the inter-rater reliability of the California Attorney's Examination is consistently reported to be below .5, there can be little doubt that the reliability of the decisions made on the basis of the scores is extremely low, and not acceptable.

10. The facts and opinions submitted in this Declaration and in the attached Evaluation of the Psychometric Adequacy of the California Attorney's Examination are true and correct to the best of my knowledge. This Declaration is submitted under penalty of perjury under state and federal law.

Dated:

July 30, 2010


PHILLIP L. ACKERMAN, Ph.D.

DECLARATION OF GARY H. McCLELLAND, Ph.D.

I, GARY H. McCLELLAND, Ph.D. am competent to testify and have personal knowledge, and based on that knowledge declare the following:

1. I'm a psychology professor at the University of Colorado at Boulder. I am interested in the study of judgment and decision making, measurement and scaling, and statistics and data analysis. I do research in these areas and teach courses about statistics and measurement.

2. I have a Ph.D. (1974) from The University of Michigan, and I am also a Faculty Fellow, at the Institute of Cognitive Science.

3. A representative of my publications includes:

- *McClelland, G.H. (2000). Seeing Statistics. Duxbury Press.*
- *Judd, C.M., & McClelland (1998). Measurement. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), The handbook of social psychology (4e). Cambridge University Press.*
- *McClelland, G.H. (1997). Optimal design in psychological research. Psychological Methods, 2, 3-19.*
- *Judd, C.M., McClelland, G.H., & Smith, E.R. (1996). Testing treatment by covariate interactions when treatment varies within subjects. Psychological Methods, 1, 366-378.*
- *Judd, C.M., McClelland, G.H., & Culhane, S.E. (1995). Data analysis: Continuing issues in the everyday analysis of psychological data. Annual Review of Psychology, 46, 433-465.*
- *McClelland, G.H., & Judd, C.M. (1993). Statistical difficulties of detecting interactions and moderator effects. Psychological Bulletin, 114(2), 376-390.*

- *McClelland, G.H., Schulze, W. D., & Coursey, D. L. (1993). Insurance for low-probability hazards: a bimodal response to unlikely events. Journal of Risk and Uncertainty, 7(1), 95-116.*
- *Irwin, J. R., Slovic, P., Lichtenstein, S., & McClelland, G. H. (1993). Preference reversals and the measurement of environmental values. Journal of Risk and Uncertainty, 6, 5-18.*
- *Judd, C. M., & McClelland, G. H. (1989). Data analysis: A model comparison approach. New York: Harcourt Brace Jovanovich.*

4. I have previously studied the Colorado bar examination from 1972 to 1975, and based on that study wrote "Assessing Bias in Professional Licensing Examinations by Checking Internal Consistency," 9 Law and Human Behavior, No. 3, p. 305 (1985).

5. I have reviewed Dr. Phillip L. Ackerman's "Evaluation of the Psychometric Adequacy of the California Attorney's Examination" dated February 15, 2008, and generally agree with it. Dr. Ackerman is a credible psychometrician as well. In my opinion, the lack of an explicit equating procedure for the Essay and Performance Test sections is a fatal flaw. The degree of inter-rater agreement is dreadful. I do not believe any scientist would ever publish data based on such low inter-rater agreement.

6. The facts and opinions submitted in this Declaration are true and correct to the best of my knowledge. This Declaration is submitted under penalty of perjury under state and federal law.

Dated: 5/24/08


GARY H. McCLELLAND, Ph.D.

Evaluation of the Psychometric Adequacy
of the California Attorney's Examination

Prepared by Phillip L. Ackerman, Ph.D.
February 15, 2008

1. According to the State Bar of California Committee of Bar Examiners/Office of Admissions. Retrieved from the web 12/8/07

“To be admitted to practice law in California, an attorney applicant must comply with the requirements outlined in the *Rules*, which include: 1) registration as an attorney applicant; 2) a positive moral character determination; 3) passage of the Multistate Professional Responsibility Examination (MPRE); and 4) passage of the California Bar Examination.”

“Attorney applicants admitted in other states or jurisdictions of the United States who have been admitted in active status in good standing four years immediately preceding the first day of the administration of the California Bar Examination, may elect to take the Attorney's Examination, which is of two days duration and consists of six essay questions and two performance test questions from the California Bar Examination.”

2. After review of the available materials, it is concluded that the Attorney's Examination fails to meet the *Standards for Educational and Psychological Testing*. (1999), Published by the American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education.

The specific shortcomings of the test are as follows:

- a. The Attorney's Examination lacks content-related validity. According to Steven P. Klein (The Rand Corporation):
“State bar examinations have been criticized for measuring only a few of the important skills and abilities that are needed for the practice of law. For example, a typical essay question provides several facts that are material to a case and then asks the applicant to determine how the case should be resolved relative to the applicant's knowledge of general legal principles. **The exam does not address interviewing, negotiating, or oral advocacy skills; the ability to draft or evaluate legal documents;**

or the ability to conduct legal research.” (Klein, 1983a; *Measuring Legal Research Skills on a Bar Examination*, Rand Report P-6879). [emphasis added]¹

Similarly, Klein claimed “Some of the other skills that are important for legal practice that are not tested directly by the traditional bar exam include the ability to interview clients, examine a witness, conduct legal research, and negotiate a settlement.” (Klein, 1983b, *Relationship of Bar Examinations to Performance Tests of Lawyering Skills*, Rand Report P-6895)

- b. The Attorney's Examination has never been demonstrated to have criterion-related validation, in terms of evaluating the scores on the test and comparing them to performance of practicing attorneys. (per Dr. Susan M. Case, Director of Testing, National Conference on Bar Examiners, 1/18/08).
- c. The scores on the Attorney's Examination are determined in a manner that is not consistent with professional standards. The reliability of the test scoring procedures fails to reach a level that would be acceptable for high-stakes testing. (Specifically, inter-rater agreement is quite low, a correlation of .48 between raters indicate only 23% shared variance among ratings; source: Klein & Bolus; Gansk & Associates 2003.) An acceptable level of reliability for such high-stakes testing would be shared variance in the neighborhood of 70% or higher (corresponding to reliability of about .84 or higher).
- d. The passing cut-off score on the Attorney's Examination is determined without regard to the actual knowledge and skills of the individuals taking the examination, but is determined by the scores of other individuals taking the test.

¹ The incomplete coverage of the content of the job in this case might be compared to a driving license examination that only involves parking a car, and does not involve driving outside of a parking lot. Such an examination would fail to adequately sample the content of the overarching activities allowed by a driver's license.

The adjustment process that is used to transform the Attorney's Examination raw scores to a mean and standard deviation that matches the results of **other** examinee's MBE scores, is performed in a manner that is inconsistent with actually assessing the individual examinee's knowledge and skills. This process is ". . . arbitrary, because it ensures that some fixed proportion of applicants will fail even though all or most of the applicants may in fact be qualified. The more able the group of applicants taking the test, the higher the passing scores will be." (Shimberg, 1981, p. 1141)

- e. Because of the lower reliability of the Attorney's Examination, when compared to the MBE, and the non-compensatory scoring for the Attorney's Examination, the resulting regression-to-the-mean (a statistical phenomena that is exacerbated by lower reliability) will result in fewer qualified individuals taking the Attorney's Examination actually obtaining a passing score on the examination, in comparison to the California Bar Examination.

Specific Evidence/Data that Form the Basis of the Evaluation

1. The described procedure (from the Gansk & Associates "Analysis of the February 2003 California Bar Examination" by S. P. Klein & R. Bolus; and repeated in subsequent reports)
 - a. The procedure described for the "equating" of test scores from the MBE (using common questions across multiple administrations of the test) appears to be an appropriate use of psychometric procedures to maintain approximately similar meanings of test scores from one administration to the next. However, neither the National Conference of Bar Examiners (phone contact with Dr. Susan Case, 1/18/08) nor ACT, Inc. (phone contact with Diane Johnson, 1/18/08) would make available any evidence regarding the adequacy of the procedures used for equating the test scores. A failure to release such information is inconsistent with standards for open professional evaluation of the psychometric adequacy of tests used for commercial purposes.
 - b. The reported reliability (form of reliability not specified in the report) of the Attorney's Examination ranges from .68 to .79 in available reports, but is always reported to be lower than the MBE (which ranges from .87 to .90) [Given that no discussion is presented about a sample of individuals who have taken the test twice for test-retest reliability purposes, I have surmised that the authors are referring to internal consistency reliability.] Because internal consistency reliability estimates (such as Cronbach's α) represent a confounding of test reliability and homogeneity (see Ackerman & Humphreys, 1991), the reported reliability information does not, in and of itself, provide sufficient evidence to determine if the test has adequate reliability. A sample of individuals who have taken the test twice, which would provide either test-retest same form or test-retest alternate form reliability is needed, in order to ascertain whether the reliability of the test and the stability of the knowledge and skills assessed by the test are adequately measured. Test-retest alternate form reliability would be the most appropriate form of reliability for a certification test, and such data are not available.

- c. No criterion-related validity data are reported for the MBE or the Attorney's Examination. According to the evidence available and statements from the National Conference of Bar Examiners (NCBE) Director of Testing, there has never been an attempt to establish whether the test has criterion-related validity.
2. There is no clear indication of how the examination questions for the Essay section and the Performance Test (PT) section were created.
3. There is no evidence of explicit equating procedure (as is performed with the MBE) which involves repeated use of a subsample of items from one administration to the next, in order to retain equivalent interpretations of test scores.
4. Instead, the raw scores for the written portion (Essay and PT) are converted to another scale, by means of multiplying the raw scores by a constant and subtracting the mean, in order to "have the same mean and standard deviation as the applicants' MBE scores" (Klein & Bolus, 2003, p. 5)

"An applicant's Total scale score was a weighted combination of that applicant's MBE and Written scale scores. The formula for computing Total scale scores was:

$$\text{Total Scale Score} = (.35)(\text{MBE Scale}) + (.65)(\text{Written Scale})" \text{ (p.3)}$$

The actual weighting and constant values applied to the raw scores on the Written portion changes from one administration to the next; essentially moving the goal posts, depending on the scores of other examinees who complete the MBE. For example, in July, 2002 the Written Scale score was "(3.0256 x Written Raw) - 473.9788" (p. 3). In February, 2003, the Written Scale score was "(3.2419 x Written Raw) - 584.0536"(p. 3)

5. As noted above, according to the State Bar of California "Attorney applicants who take the Attorney's Examination also have their scores on the written section placed on the same scale of measurement as general applicants, but as they are exempt from the MBE, their pass/fail status is based solely on the written section" (p. 3).

On the basis of professional standards:

Standards for educational and psychological testing. (1999). Published by the American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education

1. Ideally, there would be criterion-related validity for each of the components of the Bar Examination (in particular, both the MBE and the Written sections).

Comment: There do not appear to be any empirical data on criterion-related validity for either of these tests.

2. In the absence of criterion-related validity, there should be indicators of content-based validity.

Comment: It is not possible, from the information made available, to ascertain the process by which items are created for the Written sections of the test.

3. Page 157. "Tests for credentialing need to be precise in the vicinity of the passing, or cut, score. They may not need to be precise for those who clearly pass or clearly fail."

According to the Gansk & Associates 2003 report:

"There were 991 applicants who had their answers read at least twice. On the average, an applicant's total written raw score on the first reading was 10 points higher than it was on the second reading. The correlation between these scores was .48. *This value underestimates the true degree of agreement between the readers because reread was limited to applicants near the pass/fail line.*" (p.5., italics added)

In July 2002, "The correlation between these scores was .40"

In the reports of the February, 2005 and July 2005 administrations, these statistics have been omitted.

Comment: This level of inter-rater agreement is quite low, indicating only

16% to 23% shared variance among ratings. With such low reliability, there is a strong tendency for regression-to-the-mean when scores are averaged. If the pass/fail cutoff score is above the mean, the result will be lower passing rates than would be obtained if the procedure were more reliable. (This results in a larger portion of truly qualified individuals taking the Attorney's Examination to fail the examination, in comparison to individuals taking the full California Bar Examination, *ceteris paribus*.)

4. Page 157 "Legislative bodies sometimes attempt to legislate a cut score, such as a score of 70%. Arbitrary numerical specifications of cut scores are unhelpful for two reasons. First, without detailed information about the test, job requirements, and their relationship, sound standard setting is impossible. Second, without detailed information about the format of the test and the difficulty of the items, such numerical specifications have little meaning."

Page 158. "Some credentialing groups consider it necessary, as a practical matter, to adjust their criteria yearly in order to regulate the number of accredited candidates entering the profession. This questionable procedure raises serious problems for the technical quality of the test scores. Adjusting the cut score annually implies higher standards in some years than in others, which, although open and straightforward, is difficult to justify on the grounds of quality of performance. Adjusting the score scale so that a certain number or proportion reach the passing score, while less obvious, is technically inappropriate because it changes the meaning of the scores from year to year."

Standard 14.13. "When decision makers integrate information from multiple tests or integrate test and nontest information, the role played by each test in the decision process should be clearly explicated, and the use of each test or test composite should be supported by validity evidence. (p. 181).

Standard 14.17. "The level of performance required for passing a credentialing test should depend on the knowledge and skills necessary for acceptable performance in the occupation or profession and should not be adjusted to regulate the number of proportion of persons passing the test."

Comment: Given that there is no apparent matching between the content of the Written portion *scores* and the content of the knowledge and skills of the occupation (**especially given that the score 'meaning' is norm-centered, rather than content-centered, based on the transformations to the raw scores**), it seems clear that these standards have not been met in any explicit fashion.

5. On January 18, 2008, I spoke by telephone with Dr. Susan M. Case, Director of Testing, National Conference on Bar Examiners.

Dr. Case confirmed that there have been no validation studies conducted on the Multistate Bar Examination (MBE) that involve practicing lawyers. That is, there are no predictive or concurrent validity data that would support the use of the MBE for determining the competency of individuals for admission to the Bar.

Dr. Case also indicated that NCBE would not release any information about the psychometric adequacy of the MBE test (e.g., reliability and validity).

I also spoke with Diane Johnson of ACT on January 18, 2008. She noted that ACT develops and equates the test for NCBE, but she indicated that no information could be released about the psychometric properties (e.g., reliability, equating) could be released from her organization.

References

Klein, S. P., & Bolus, R. (2002). *Analysis of the July 2002 California Bar Examination*. This is an attachment to the Report to the Supreme Court of California.

Klein, S. P., & Bolus, R. (2003). *Analysis of the February 2003 California Bar Examination*. This is an attachment to the Report to the Supreme Court of California.

Klein, S. P., & Bolus, R. (2005a). *Analysis of the February 2005 California Bar Examination*. This is an attachment to the Report to the Supreme Court of California.

Klein, S. P., & Bolus, R. (2005b). *Analysis of the July 2005 California Bar Examination*. This is an attachment to the Report to the Supreme Court of California.

Report to the Supreme Court of California on the July 2002 California Bar Examination

Report to the Supreme Court of California on the February 2003 California Bar Examination

Report to the Supreme Court of California on the February 2005 California Bar Examination

Report to the Supreme Court of California on the July 2005 California Bar Examination

Standards for educational and psychological testing. (1999). Washington DC: American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education.

Ackerman, P. L., & Humphreys, L. G. (1991). Individual differences theory in industrial and organizational psychology. Chapter in M. D. Dunnette & L. M. Hough (Eds.) *Handbook of industrial and organizational psychology* (Volume 1, pp. 223 - 282). Palo Alto, CA: Consulting Psychologists Press.

Klein, S. P. (1983a); *Measuring legal research skills on a bar examination*, Rand Report P-6879. Rand Corporation.

Klein, S. P. (1983b). *Relationship of bar examinations to performance tests of lawyering skills*. Rand Report P-6895. Rand Corporation

Shimberg, B. (1981). Testing for licensure and certification. *American Psychologist*, 36, 1138-1146.

Additional Citations and Reference Materials

Others have offered both psychometric and principled arguments against the methods currently in use for deriving scores on the Written Section (Essay and PT) of the Bar Examination. For example, see the especially detailed analysis provided by:

Merritt, D. J., Hargens, L. L., & Reskin, B. F. (2001). Raising the bar: A social science critique of recent increases to passing scores on the bar exam. *University of Cincinnati Law Review*, Vol. 69, No. 2.

For a brief review of these points, see pp. 9-12 by D. J. Merritt in *The Bar Examiner*, November 2001.

These issues are also discussed in W. C. Kidder (2004). The bar examination and the dream deferred: A critical analysis of the MBE, social closure, and racial and ethnic stratification. *Law and Social Inquiry*, pp. 547-589.