

OPEN SESSION

AGENDA ITEM:

BOARD OF TRUSTEES MEETING – MAY 2018

PROGRAMS COMMITTEE

ITEM III.G.

DATE: May 8, 2018

TO: **Members, Programs Committee**

FROM: Lisa Jeong Cummins, Program Manager III, Examinations
Amy C. Nuñez, Interim Director III, Admissions

SUBJECT: Psychometrician Report on Two-Day Exam Results

EXECUTIVE SUMMARY

In March of 2015, the Committee of Bar Examiners decided to adopt a modified format for the California Bar Examination, converting it from a 3-day examination to a 2-day examination and changing the weighting of its component written and multiple-choice parts. The Board of Trustees unanimously approved the proposal in July of 2015 and the California Supreme Court approved the change in March of 2016. The July 2017 administration of the California Bar Examination was the first time using the modified format. The Committee of Bar Examiners requested its psychometrician to conduct an analysis and prepare a report, the results of which are summarized below.

BACKGROUND

The first administration of the modified California Bar Examination (2-Day General Bar Examination and 1-Day Attorneys' Examination) took place in July 2017.

The Committee of Bar Examiners (Committee) requested its psychometrician, Roger Bolus, Ph.D., to conduct an analysis and prepare a report on the July California Bar Examination (CBX) to address three core questions:

- (1) To what degree have the scores and subsequent bar passage rates been impacted by the modified CBX?
- (2) To what degree have differences in scores and subsequent passage rates between key demographic subgroups been impacted by the modified CBX?
- (3) To what degree was the reliability of the CBX (and its components) impacted by shortening the examination and modifying the weighting of the respective sections?

Dr. Bolus' report is attached to this agenda item.

DISCUSSION

The key findings emerging from Dr. Bolus' report can be summarized as follows:

- The modifications made to the bar examination had no differential impacts on the component scores, total test scores or passage rates of women versus men or white versus minority applicants.
- The overall reliability of the examination was not negatively impacted under the modified examination structure, and in fact improved, aided in part by recent increases in the reliability of the Multistate Bar Examination (MBE) and written sections.
- Under the new weighting scheme (50% written and 50% multiple-choice) applied to the modified examination format, only 0.7% fewer applicants passed the examination that would have otherwise passed under the previous 65/35 % weighting scheme. All of those applicants did, however, go into regrade and eventually passed.
- For the first time since 2008, CBX scores and passage rates reversed their downward trend in 2017. The July 2017 bar passage rate increased by 6% over the July 2016 examination.
- A separate analysis of those repeating the CBX for the first time during each of the two study timeframes (July 2015/July 2016 and July 2016/July 2017) was also conducted. This analysis revealed that the group of repeaters first taking the 3-day examination in July 2016 and then repeating it in July 2017 under the new 2-day format exhibited statistically significant larger increases in their scores and passage rates when compared to those who took the 3-day version for both their first test cycle in July 2015 and again in July 2016.

Dr. Bolus concluded that implementation of the modified examination format for the CBX did not have any discriminatory impacts on applicants or negative effects on test reliability.

Definitive conclusions, however, could not be easily drawn regarding the reported increases in test scores and passing rates observed with the July 2017 CBX. With regard to the theory that the modified CBX was somehow "easier" than its predecessor, Dr. Bolus opines that the parallel improvement in the MBE scores nationally in 2017, following a similar period of decline as observed in California, argues against this interpretation. Additional analyses also established that the change to the weighting scheme under the modified examination was not a factor. Other possible contributing factors, including differences in applicants' abilities and/or preparation, and the statistical phenomenon of regression toward the mean, are discussed in the report. Although Dr. Bolus cannot provide a definitive explanation of the observed increases in scores, on the balance, he found nothing that pointed to the change in examination format as the primary cause of these improvements.

FISCAL/PERSONNEL IMPACT

None.

RULE AMENDMENTS

None.

BOARD BOOK AMENDMENTS

None.

STRATEGIC PLAN GOALS & OBJECTIVES

Goal: 2. Ensure a timely, fair, and appropriately resourced admissions, discipline, and regulatory system for the more than 250,000 lawyers licensed in California.

Objective: After the results of the February 2019 Bar Exam are published, evaluate the results of the two-day exam on pass rates and costs.

RECOMMENDATION

It is recommended that the Programs Committee and the Board of Trustees approve the following resolution:

RESOLVED that the report on Analysis of the First Two-Day Administration of the California Bar Examination, prepared by Roger Bolus, Ph.D. and dated March 2, 2018, be received and filed.

ATTACHMENT LIST

- A.** Analysis of the First Two-Day Administration of the California Bar Examination

**Analysis of the First Two-Day Administration
of the California Bar Examination**

Roger E. Bolus, Ph.D.

Research Solutions Group

March 2, 2018

A report prepared for the California Committee of Bar Examiners

EXECUTIVE SUMMARY

In 2015, the Committee of Bar Examiners (CBE) recommended that the California Bar Examination (CBX) be modified so that administration of the General Bar Examination could be shortened from a 3-day to 2-day period and the Attorneys' Examination could be shortened from a 2-day to a 1-day period¹. This recommendation was supported by the outcomes of statistical simulations which indicated that reducing the number of written and performance tasks included in the exam while also increasing the statistical weightings assigned to the MBE would not impact either the reliability or validity of the test.

Following the California State Supreme Court's approval, the modified CBX was used for the first time for the July 2017 administration of the exam. The CBE requested an analysis of the results of this initial implementation of the 2-day administration to address three core questions: (1) to what degree were the scores and subsequent bar passage rates impacted by the modified CBX? (2) to what degree were differences in scores and subsequent passage rates between key demographic subgroups impacted? and (3) to what degree was the reliability of the CBX (and its components) impacted?

The key findings emerging from this evaluation include the following:

- The modifications made to the bar examination had no differential impacts on the component scores, total test scores or passage rates of women versus men or white versus minority applicants, with the single exception of the average Written Scale Score for Asians. The score for Asian applicants did fall slightly in 2017 (to 1,394 from 1,411, and 1,396 in 2015 and 2016, respectively) under the new test format while all other racial/ethnic groups saw actual increases. The drop, however, was not large enough to lead to a significant impact on either Total Scale Scores or bar passage rates. These findings substantiate the outcomes of the earlier simulations conducted prior to modification of the exam which predicted that passing rates within the

¹ The California Bar Examination, as it was formulated prior to July 2017, was comprised of the 3-day General Bar Examination and a 2-day Examination for attorneys. The Attorneys' Examination is the same test as the General Bar Examination, but without the Multistate Bar Examination (MBE) component. For ease of reference throughout this report, the California Bar Examination (CBX) and the General Bar Examination shall be used interchangeably. Data from the Attorneys' Examination was not considered in the analysis for this report.

respective gender and racial/ethnic groups would only be impacted by 1% to 2%, subsequently resulting in virtually no change in the relative passing rates of different genders of racial/ethnic groups.

- As also predicted in the earlier simulations, the overall reliability of the examination was not negatively impacted under the modified examination structure. Indeed, the test reliability actually improved, aided in part by recent increases in the reliability of the MBE and written sections. The reliability of the July 2017 examination was .92, the highest of the three years examined, and the highest reported in recent history. This high level of reliability is a function of the .93 reliability of the MBE, (which has risen fairly steadily over the previous 3 to 4 years and plateaued around .92 to .93) and the 50% weighting given to the MBE in the calculation of the Total Score.
- Under the new 50/50 weighting scheme applied in the 2-day exam format, only .7% fewer applicants (n=198) passed the exam that would have otherwise passed under the previous 65/35 scheme. All of these 198 applicants went into regrade and eventually passed, however.
- For the first time since 2008, CBX scores and passage rates reversed their downward trend in 2017. MBE scores rose by 9 points to 1432 in 2017, after reaching their lowest level in 2016 (1423). The improvement in the MBE was also observed nationwide, where scores increased by 14 points from 1403 to 1417. For California applicants, the scores for the Written Section, which is scaled to the MBE, and the Total Scale Scores also increased from previous years, and the July 2017 bar passage rate increased by 6% over the July 2016 examination. Analyses of the overall improvements that were observed could not be directly linked to changes in the examination structure.
- A separate analysis of those repeating the bar for the first time during each of the two study timeframes (July 2015/July 2016 and July 2016/July17) was also conducted. This analysis

revealed that the group of repeaters first taking the 3-day test in July 2016 and then repeating it in July 2017 under the new 2-day format exhibited statistically significant larger increases in their scores and passage rates when compared to those who took the 3-day version for both their first test cycle in July 2015 and again in July 2016. This is an intriguing finding since the two groups showed no statistically significant differences in their MBE scores on their first testing, suggesting that the two cohorts were of roughly equivalent ability.

The primary conclusion drawn from the first three of these findings is that, as predicted in the original simulations conducted to predict the outcomes of a shortened examination, implementation of a modified 2-day exam format for the CBX did not have any discriminatory impacts on applicants or negative effects on test reliability. These outcomes were achieved while reducing both the testing burden on applicants and the costs associated with administering the examination.

Definitive conclusions regarding the reported increases in test scores and passing rates observed with the July 2017 implementation of the new 2-day format are not as easily drawn however. The parallel improvement in MBE scores nationally in 2017, which had exhibited a similar period of decline as observed in California, argues against the propositions that the modified test format was “easier”. Given the limitations of data on possible causes, however, we cannot provide a definitive explanation of the observed increases in scores. On the balance, however, we found nothing that pointed to the change in exam format as the primary cause of these score improvements. Further studies on subsequent administrations are warranted.

INTRODUCTION & BACKGROUND

The Standards for Educational and Psychological Testing, the de facto set of guidelines for license testing, contains multiple standards that directs testing agencies to evaluate the impact that fatigue plays on examinee performance and directs those organizations to continually consider methods for improving efficiencies in test development and administration, while maintaining psychometric quality. Up until 2017, California was one of only five states in which the bar examination exceeded two days duration. The California Bar Examination (CBX) was given over a 3-day period and included six 1-hour essay questions, two 3-hour performance tasks and the 200-item Multistate Bar Examination (MBE) multiple choice test.

Over the past several years, the California Committee of Bar Examiners has focused on assessing the feasibility of shortening the examination. A 2011 study sponsored by the Committee (Klein & Bolus, 2011) evaluated multiple alternative configurations for the exam created by reducing the number of written questions and modifying the statistical weightings applied to the different sections of the test. Based on simulations using data from 20 separate administrations between 2001 and 2010, the study identified several alternatives that could maintain *test reliability* while limiting testing to two days.² Subsequent studies of the impact of alternative test formats on *test validity* (Buckendahl, 2013) were also conducted.

Based on research findings, the Committee of Bar Examiners recommended in 2015 that the California Bar Examination be modified to include five 1-hour essay questions, a single 90-minute performance task, and the MBE. Test administration would be shortened to a two-day period from the traditional three days and the statistical weightings assigned to each section of the exam would change. Formal notice of the impending change was given to the public in July of 2015, with a formal request for a July 2017 adoption going to the State Supreme Court in March 2016 (Holton & Grunberg, 2016). Following the Court's approval, the modified CBX was used for the first time during the July 2017 administration of the exam.

² Follow up reports by Bolus, 2017a and 2017b replicated those findings using examination data from 2012 through February 2017.

Following the decision to change the examination, the Committee requested that results of the initial implementation of the 2-day administration of the CBX be evaluated. The primary questions to be investigated were:

1. To what degree have the scores and subsequent bar passage rates been impacted by the modified CBX?
2. To what degree have differences in scores and subsequent passage rates between key demographic subgroups been impacted by the modified CBX?
3. To what degree was the reliability of the CBX (and its components) impacted by shortening the examination and modifying the weighting of the respective sections?

This report presents the outcomes of this evaluation. The **Methodology** section discusses the data and approach used for the analyses; the **Results** section includes detailed findings relevant to each of three research questions; and the **Summary and Conclusions** section presents a summation and interpretation of the key study findings.

METHODOLOGY

In order to answer the three questions of interest, score data and applicant demographics from the most recent three July administrations of the CBX (2015, 2016 and 2017) were compiled and analyzed. The analyses focused on year-over-year comparisons contrasting shifts between the two contiguous periods where the examination had a similar 3-day structure (2015 and 2016), and the period during which the exam structure changed (2016 and 2017).

Specific analyses included:

- Replication of the simulations originally conducted to assess the impact of the proposed changes in the weighting of different components of the exam to assess the actual impact of the modified weighting scheme.
- Analysis of the performance of applicants repeating the examination during each two-year period to provide further insight into the impact of a change in the exam structure upon applicants hypothetically of the same ability level
- Multi-variate analysis to assess if the alternative exam structures had differential impacts on males vs. females and racial/ethnic minorities versus the majority; and
- Calculation of the score reliabilities for each section of the examination along with the overall score for the July 2015, 2016 and 2017 administrations.

Details and outcomes of these analyses are discussed in the Results section.

Applicant and Score Data.

Table 1 presents the counts and characteristics of applicants from the 2015, 2016 and 2017 CBX administrations used for our analyses for review. Only applicants with a full set of scored written questions and an MBE score were included. Both first time takers and repeaters were included except where noted in the discussion of a specific analysis. For specific subgroup analyses, applicants with missing demographic information (i.e., gender or race) were excluded. For gender, less than .1% of cases lacked a gender code. For analyses focusing on race, the four, major racial/ethnic groups (Asian,

Table 1

Applicant Counts and Characteristics for the July Administration

of the California Bar Examination

2015 -2017

By Sex, Race and Overall

	Year																	
	2015						2016						2017					
	1st Timer		Repeater		Total		1st Timer		Repeater		All		1st Timer		Repeater		All	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Sex																		
Female	2,901	50%	1,156	47%	4,057	49%	2,657	52%	1,226	49%	3,883	51%	2,865	53%	1,568	50%	4,433	52%
Male	2,837	49%	1,285	53%	4,122	50%	2,410	47%	1,284	51%	3,694	48%	2,462	46%	1,570	50%	4,032	47%
Missing	55	1%	2	0%	57	1%	58	1%	13	1%	71	1%	70	1%	11	0%	81	1%
Race																		
Asian	1,124	19%	568	23%	1,692	21%	999	19%	569	23%	1,568	21%	1,137	21%	716	23%	1,853	22%
Hispanic	709	12%	404	17%	1,113	14%	667	13%	445	18%	1,112	15%	711	13%	529	17%	1,240	15%
Black	271	5%	250	10%	521	6%	236	5%	264	10%	500	7%	286	5%	321	10%	607	7%
White	3,233	56%	1,051	43%	4,284	52%	2,829	55%	1,077	43%	3,906	51%	2,838	53%	1,356	43%	4,194	49%
Other	456	8%	170	7%	626	8%	394	8%	168	7%	562	7%	425	8%	227	7%	652	8%
Total	5,793	70%	2,443	30%	8,236	100%	5,125	67%	2,523	33%	7,648	100%	5,397	63%	3,149	37%	8,546	100%

Hispanic, Black and White) made up approximately 92.5% of all test takers, with the remaining 7.5% were cases either missing racial/ethnic information or having unclassifiable codes. For both gender and race, the rate of missing data was similar for each administration. The total number of applicant cases across the three years was 24,429, with 2017 including the largest number of test takers (8,546). The percentage of repeat examinees (37%) was also the highest in the same year. Demographically, the trend continued for females to make up a greater percentage of test takers rising 1% a year. And, for the first time, racial/ethnic minorities made up the majority of examinees on the CBX with Whites accounting for less than 50% in 2017. Due in part to the low passage rates in the preceding years, applicants repeating the examination in 2017 comprised more than 1 out of every 3 test takers. This is the largest rate in recent history.

RESULTS

As previously stated, the study analyses focused on year-over-year comparisons contrasting shifts between the two contiguous periods when the examination had a similar 3-day structure (2015 and 2016), and the period during which the exam structure changed (2016 and 2017). The three research questions provide the framework for the discussion of findings.

Research Question 1. To what degree have scores and bar passage rates been impacted by the modified CBX?

The data required to begin to address this question are presented in Tables 2 and 3. Table 2 reports the scale score descriptive statistics for the total population of examinees, along with the overall bar passage rates, for each of the 3 years included in the study. Since the proportion of repeat versus first-time test takers is known to impact the average test scores in a single administration, the same data are reported for first-time applicants only in Table 3.

Table 2

Descriptive Scale Score Statistics & Bar Passage Rates

All July Examinees

2015-2017

		<u>Year</u>			<u>Differences</u>	
<u>Metric</u>		<u>2015</u>	<u>2016</u>	<u>2017</u>	<u>2015- 2016</u>	<u>2016- 2017</u>
Scale MBE	Ave.	1,426	1,423	1,432	-3	9
	S.d.	158	167	168	9	1
Scale Written	Ave.	1,427	1,415	1,427	-12	12
	S.d.	155	165	165	10	0
Scale Total	Ave.	1,427	1,418	1,430	-9	12
	S.d.	145	155	154	10	-1
Passing	%	47%	44%	50%	-3%	6%

Table 3
Descriptive Scale Score Statistics & Bar Passage Rates

First-Time Test Takers Only

2015-2017

		<u>Year</u>			<u>Differences</u>	
<u>Metric</u>		<u>2015</u>	<u>2016</u>	<u>2017</u>	<u>2015- 2016</u>	<u>2016- 2017</u>
Scale MBE	Ave.	1,461	1,458	1,468	-3	10
	S.d.	155	167	170	12	3
Scale Written	Ave.	1,469	1,461	1,473	-8	12
	S.d.	152	167	167	15	0
Scale Total	Ave.	1,466	1,460	1,470	-6	10
	S.d.	142	157	157	15	0

Inspection of Table 2 shows that:

- The downward trend in CBX scores that began in 2008 continued in 2015 and 2016, but then reversed in 2017. MBE scores rose by 9 points to 1432 in 2017, after reaching their lowest level in 2016 (1423). The scores for the Written Section, which are scaled to the MBE, and the Total Scale Scores followed a similar pattern.
- The 3% drop in passing rates between 2015 and 2016 (47% to 44%) reversed to a 6% increase between 2016 and 2017 (44% to 50%).
- The standard deviations, which provide a measure of the spread in scores, increased for both sections of the exam and the total scale score between 2015 and 2016, but then remained almost identical for 2016 and 2017.

Finally, inspection of Table 3 show that the observed differences in year-over-year performance for 1st time test takers were roughly equivalent to those for the applicant population as a whole.

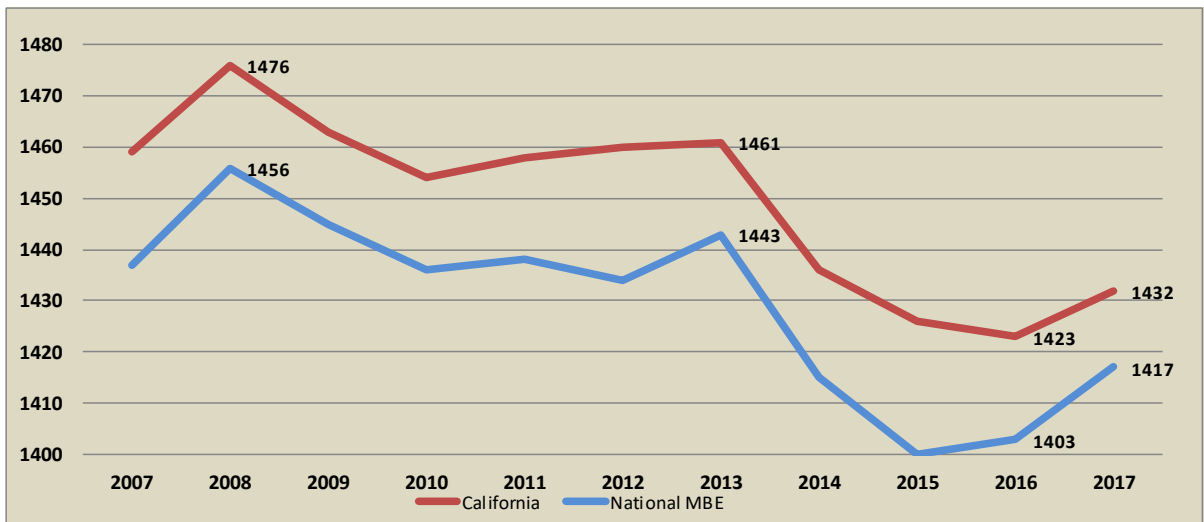
Research Question 1. Was the 2-day examination “easier”?

This ancillary question emerges immediately from an inspection of Tables 2 and 3 since both test scores and the passage rate improved following implementation of the 2-day test format in 2017. This improvement was observed for the overall population of test takers as well as for first-time takers. Interpretation of this upturn cannot be made, however, independent of other considerations.

The MBE. As can be seen from the tables, the average MBE score increased from 2016 to 2017, following steady decreases since 2008. Several factors related to the MBE would tend to argue against the impact that the shortened exam would have had on that increase. First, the MBE was administered on the second day of testing, just as under the 3-day format. It can reasonably be assumed that applicants were no more or less fatigued while sitting for that portion of the examination under the 2-day format.

Previous analyses have also demonstrated the close relationship that has existed between national performance on the MBE and performance by California applicants. Figure 1 below illustrates the historic average performance of all U.S. MBE takers and California takers.

Figure 1
Comparison of Historic Average Performance on the July MBE
All U.S. vs. California



The graph illustrates the strong relationship that has historically existed between national and state-level performance on the MBE ($r=.97$). It also shows that, as expected from those historical results,

the increase observed in California in 2017 was also observed nationally (where no state had changed its examination format from 2016 to 2017).³

Furthermore, potential changes in applicant ability and/or preparation should also be considered in evaluating the 2017 bar results. Every administration of the MBE is equated to previous anchor exams to ensure that any changes in performance over time on the MBE are a function of applicant ability rather than differences in the test itself. Unfortunately, at this point we do not have access to independent measures of ability (e.g., entering law school credentials and law school performance, which are both highly correlated with MBE performance) for the test takers included in this study. As such, we cannot investigate the extent to which applicant ability impacted test performance during this period. We also understand through anecdotal evidence that law schools have been intensifying their bar preparation courses in recent years; again, we presently do not have the data required to assess the extent to which such preparation might be associated with the observed improvement in test scores.

Change in Section Weighting. To maintain the historic high levels of examination reliability with a shortened written section, it was necessary to weight the MBE higher in the 2-day format (.50) than in the 3-day format (.35). Some critics suggested that this change would favor test takers that perform better on multiple choice tests and net out to an overall increase in passage rates. The simulations and modeling that were conducted on the historical exams predicted that there would be no effect on passage rates due to the weighting change.

For the sake of completeness, we replicated that analysis here, but in reverse by weighting the shortened written section at .65 and the MBE at .35. Using the original written score scaling parameters, we re-calculated a total score and compared pass/fail rates based on Phase 1 scores only. The analysis showed that .7% fewer applicants (n=198) passed under the 50/50 weighting scheme than under the previous 65/35 weighting (47.3% vs. 48.0%) after Phase 1. However, all 198 cases went into regrade and eventually passed, suggesting that section weighting changes did not make the examination any more or less difficult.

Performance of Repeating Applicants. Another insight into the impact of the modified CBX on applicants' scores can be gained by studying the results of test repeaters. First-time test takers in 2015 who failed and repeated in 2016 took the examination under a common 3-day format. The

³ A regression equation predicting CA July 2017 from historic national results suggested that CA performance should have been slightly higher than it actually was (1438 vs 1432).

group of failing first time test takers in 2016 retaking the examination in July 2017, however, took the modified CBX. Both groups of applicants could be hypothesized to be of roughly equal ability in that they both scored < 1440 on their first attempt. As such, any differences between the two groups in the change in performance might be attributed in part to the different exam structure.

To examine this possibility, first-time takers who failed in either 2015 or 2016, and repeated the following July examination were culled from the database. Their scores on their initial and subsequent examinations were linked, and the difference scores between each attempt were calculated along with their pass/fail status on the repeated exam. Statistical comparisons were then made between the two groups.

There were 742 first-time test takers in July 2015 who failed and retook the July 2016 examination, and 807 first-time test takers in July 2016 who failed and retook the July 2017 exam. A comparison of the initial MBE scores of the 2015 Group (Ave.=1,294, Sd.=122) and 2016 Group (Ave.=1,286, Sd.=125), indicated no statistical significant differences ($p=.178$) suggesting that the two cohorts were of roughly equivalent ability. Table 4 presents the differences in the average scores of each group upon re-testing and a comparison of those differences between the two cohorts.

Table 4

Comparison of the Differences in Scores

For Applicants* Repeating the Examination

2015-2017

<u>Metric</u>	<u>2 Year Period</u>		<u>Differences</u>
	<u>2015 to 2016</u>	<u>2016 to 2017</u>	
MBE	+81	+101	20
Written	+41	+93	53
Total	+55	+95	40
% Passing	24%	39%	15%

* Includes only applicants failing in their first attempt in July of the baseline year (2015 or 2016) and repeating the exam the following July (2015 or 2016)

The results from Table 4 indicate repeating applicants from both the 2015 and 2016 July examination cohorts improved upon repeating the examination in the subsequent year. However, the performance improvement in the 2016 cohort taking the 2-day format when repeating was statistically higher ($p < .001$) than for the 2015 cohort where the applicants repeated the exam using the prior 3-day format. Additionally, while 24% of the 2015 cohort passed on their subsequent attempt, fully 39% of the repeating 2016 cohort passed, a 15% difference ($p < .001$).

While these findings could suggest that applicants repeating the bar exam under a shortened format may find the test easier, there again are alternative explanations to consider. First, MBE scores increased for the overall population of test takers in 2017 in keeping with the national trend. It is also possible that exam preparation was intensified, and that the rising tide lifted all boats (i.e., both first-timers and repeaters). Finally, it might also be a possibility that these were aberrant years. That is to say, the 2015 to 2016 increases in repeaters' scores may have been lower than in previous testing cycles, and/or the increases in the 2016 to 2017 scores for the cohort of repeaters may have been particularly high.⁴

The degree of improvement in the performance of first time test repeaters observed during the first implementation of the modified CBX format in July 2017 as compared to the prior year remains notable, however. Analyses of repeaters' performance in future testing cycles will be required to establish if this remains a consistent pattern or if the magnitude of the differences varies.

Research Question 2. To what degree have the scores and subsequent passage rates for key demographic subgroups been impacted by a modified CBX?

Table 5 presents the average performance and bar passage rates for the July 2015, 2016 and 2017 examinations for males and females.

⁴ A subsequent analysis looking at the cohort of July 2014 takers that repeated the exam in 2015, revealed significant differences in the relative ability level (as measured by initial MBE) as compared to the 2016 repeating cohort. This may suggest that the large improvements observed in the 2016-2017 cohort may have been some form of regression effect for repeaters.

Table 5
Descriptive Scale Score Statistics & Bar Passage Rates
For Male and Female Test-Takers

2015-2017

Metric	Female			Male			Difference (F-M)		
	2015	2016	2017	2015	2016	2017	2015	2016	2017
MBE	1,411	1,403	1,417	1,442	1,443	1,448	-31	-40	-31
Written	1,438	1,421	1,435	1,417	1,409	1,419	21	12	16
Total	1,428	1,415	1,426	1,425	1,421	1,434	3	-6	-8
% Pass	<u>48%</u>	<u>43%</u>	<u>49%</u>	<u>46%</u>	<u>44%</u>	<u>50%</u>	<u>2%</u>	<u>-1%</u>	<u>-1%</u>

Inspection of Table 5 reveals that historical differences observed between male and female performance on the examination persisted under the 2-day exam format. Men continued to perform higher on the MBE (1,448 vs. 1,417) while women performed higher on the Written Section (1,435 vs. 1,419). This was not the case for the Total Scale Score or the passage rates. With respect to the Total Scores, under the differential weighting scenario (i.e., 35/65) used in 2015 and 2016, overall Total Scale Scores were roughly equivalent for men and women, varying on average by only +3 and -6 points between them. Under the 50/50 weighting scenario used in 2017, the overall difference between men and women was only 2 scale score points more in 2017 than in 2016. All of this resulted in only a 1% difference in passing rates between men and women in 2017, which was exactly the size of the difference observed in 2016.

To further evaluate the statistical significance of these differences, a multivariate model was applied to each metric. The model evaluated three factors, exam year, gender, and the interaction of gender by year. If the latter factor (gender by year) was found to be significant, it would imply that the size of the difference between men and women either grew or got smaller with each administration. If this were the case, it would lend evidence that the change in weighting and/or format had an adverse

impact on one of the gender groups. A non-significant finding would suggest that any sized difference between men's and women's performance in 2017 vs. previous years was due to chance alone⁵.

As was expected, the modeling identified a statistically significant difference significant ($p < .001$) in the MBE and Written scores of men and women in each of the three years. *None of the models, however, identified a statistically significant (at $\alpha = .05$) interaction of gender by year. This outcome indicated that the modifications made to the bar examination in 2017 had no differential impact on either groups' MBE, Written, or Total Scores or on passage rates.*

Furthermore, the three factors (year, gender and the interaction) included in the model accounted for less than 1.1% of the variation in applicant scores. This suggests that the observed score differences over the past three years and the male vs. female effect contribute little to understanding variations in examination performance.

Tables 6 and 7 present the same set of statistics broken down by the four, primary racial/ethnic groups. Table 6 presents the statistics themselves, while Table 7 shows the year-to-year differences for each minority group compared to the White test-taking group.

Table 6
Descriptive Scale Score Statistics & Bar Passage Rates
By Racial/Ethnic Group
2015-2017

	<u>Asian</u>			<u>Black</u>			<u>Hispanic</u>			<u>White</u>		
<u>Metric</u>	<u>2015</u>	<u>2016</u>	<u>2017</u>	<u>2015</u>	<u>2016</u>	<u>2017</u>	<u>2015</u>	<u>2016</u>	<u>2017</u>	<u>2015</u>	<u>2016</u>	<u>2017</u>
MBE	1,398	1,383	1,400	1,364	1,360	1,371	1,389	1,388	1,401	1,458	1,459	1,465
Written	1,411	1,396	1,394	1,348	1,327	1,357	1,397	1,381	1,404	1,452	1,445	1,458
Total	1,406	1,391	1,397	1,354	1,338	1,364	1,394	1,384	1,403	1,455	1,450	1,462
% Pass	<u>42%</u>	<u>38%</u>	<u>44%</u>	<u>27%</u>	<u>21%</u>	<u>30%</u>	<u>37%</u>	<u>34%</u>	<u>40%</u>	<u>55%</u>	<u>52%</u>	<u>57%</u>

⁵ Because of the size of the annual test taking population (> 7,500 in each year) in any statistical analysis there would be a tendency to identify even the smallest of differences as significant. As such, the absence of any statistically significant interactions in this case would lend even more confidence that there was no effect from the change in the exam format.

Table 6 shows that for all racial/ethnic groups, there was a decrease in scores and passage rates between 2015 and 2016 and a rise between 2016 and 2017 that parallels the pattern observed for the entire pool of applicants together in each year respectively. (See Table 2). For example, the passing rate among Asians was 42% in 2015, dropped to 38% in 2016, and then increased to 44% in 2017. Similarly, Whites passed at a rate of 55% in 2015, fell to a passing rate of 52% in 2016 and increased to 57% in 2017.

For easier reference, Table 7 presents the differences between the scores and passage rates for Whites and each minority group by year.

Table 7
Differences Between Average Scores & Bar Passage Rates
For Minority Groups When Compared to Whites
2015-2017

	<u>Asian</u>			<u>Black</u>			<u>Hispanic</u>		
Metric	<u>2015</u>	<u>2016</u>	<u>2017</u>	<u>2015</u>	<u>2016</u>	<u>2017</u>	<u>2015</u>	<u>2016</u>	<u>2017</u>
MBE	-60	-76	-65	-94	-99	-94	-69	-71	-64
Written	-41	-49	-64	-104	-118	-101	-55	-64	-54
Total	-49	-59	-65	-101	-112	-98	-61	-66	-59
% Pass	<u>-13%</u>	<u>-14%</u>	<u>-13%</u>	<u>-28%</u>	<u>-31%</u>	<u>-27%</u>	<u>-18%</u>	<u>-18%</u>	<u>-17%</u>

Similar to the analysis of gender interactions, a series of multivariate models were created to examine racial/ethnic effects. Each model evaluated three factors: exam year, race/ethnicity, and the interaction of race/ethnicity by year. Again, if the race by year interaction was found to be significant, it would imply that the size of the difference between the various races changed with each administration. A significant interaction would suggest that the change in weighting and/or format for the 2017 2-day CBX had a differential impact on one or the minority or majority groups.

Each of the models identified a statistically significant impact due to year and race independently ($\alpha=.05$), indicating that: (a) across all groups, scores and pass/fail rates were statistically different in each year; and (b) performance of the different racial ethnic groups were statistically different. However, *3 of the 4 models identified no statistically significant (at $\alpha=.05$) interaction of race by year suggesting that the modifications made to the bar examination in 2017 had no differential impact on any single group's MBE, Total Scores or passage rates.* The single exception was in the Written Score models. The source of the statistical interaction, though quite small, can be traced to Asians. In this group, Asians average Written Scale Score actually fell in 2017 (to 1,394 from 1,411 and 1,396 in 2015 and 2016, respectively) while all other racial/ethnic groups saw increases. However, the drop was not large enough to lead to have a significant impact on either Total Scale Scores or bar passage rates. The models that included racial/ethnic effects accounted for between 3% and 5% of overall score variation, as compared to slightly more than 1% in the gender models.

These results corroborate the findings of the original analyses used to provide evidence for moving to the 2-day examination and the modified weighting scheme. Those prior analyses predicted that passing rates within the respective gender and racial/ethnic groups would only be impacted by 1% to 2%, subsequently resulting in virtually no change in the relative passing rates of different genders or racial/ethnic groups.

Research Question 3. To what degree was the reliability of the CBX (and its components) impacted by shortening the examination and modifying the weighting of the respective sections?

The reliability of an examination such as the bar is impacted by three primary factors: (1) the reliability of the respective sections making up the examination, (2) the correlation between those sections and (3) the amount of weight that is given to each section to derive a final, composite score. It is well-established that the reliability of an effectively constructed multiple-choice exams with a large number of items such as the MBE tend to be highly reliable. Generally speaking, such tests are more reliable than constructed response tests, such as the Essay and Performance Test, because: (a) a multiple-choice test can sample many more areas of knowledge and skill per unit testing time than a constructed response test; and (b) the scores on a constructed response test have an added source of error, i.e., differences in grader standards and practices.

In terms of reliability, historically the written section could be given a higher weighting than the MBE because there were 8 questions. In arriving at the decision to shorten the CBX to two days, it was necessary to reduce the number of written questions. The only way that a high level of reliability could be maintained for the 2-day test format was to lower the weighting of the reduced written section and increase the weighting of the MBE. The prior simulation modeling done in studying the potential impact of a two-day examination showed that equal weighting of each section could not only maintain overall test reliability but might actually improve it.

As the July 2017 examination was the first live implementation of this test design, it was critical to evaluate whether the results of the prior simulations would actually be realized in practice. Towards this end, we calculated the reliability of each examination section, along with the Total Score, and compared it to results from both 2015 and 2016, as well as the projected range from the simulation exercises. The results are summarized in Table 8.

Table 8

Section-Specific and Total Score Reliabilities

For July 2015 – 2017 CA Bar Examinations

	<u>2015</u>	<u>2016</u>	<u>2017</u>
<u>Score</u>			
MBE	.92	.93	.93
Written	.83	.82	.79
Total	.91	.90	.92
<u>Correlation</u> <u>(MBE, Written)</u>	.70	.73	.72

Table 8 presents the score reliabilities for each section of the examination along with the overall score for the July 2015, 2016 and 2017 administrations. The reliability of the July 2017 examination was .92, the highest of the three years, and the highest reported in recent history. This high level of reliability is a function of the .93 reliability of the MBE, (which has risen fairly steadily over the previous 3 to 4 years and plateaued around .92 to .93) and the 50% weighting given to that section in the

calculation of the Total Score. Interestingly, the reliability of the shortened Written Section in 2017 (.79) had only a minor drop from the longer versions used in 2015 and 2016 (.83 and .82, respectively). The .72 correlation between the section scores under the 2-day format was fairly consistent with the correlations seen under the 3-day format in 2015 and 2016.

The differences in reliabilities between 2017 and 2015, 2016 are fairly close to what would have been predicted from the original simulations. The average Total Score reliabilities under a modified 2-day structure was estimated to be about .02 points higher than the actual exam reliabilities during 2001 and 2010. This difference approximates what is observed when comparing 2017 Total Score reliabilities to that of 2015 and 2016. The average projected Written Score reliabilities of a shortened examination during the study was (.69 +/- .04), while actual written score reliability of the 6 essay/2 PT test during that period averaged .77. With improved questions and calibration procedures, in recent years (2011 to 2016), the actual reliability of the 6 essay/2 PT written section improved to an average of .81. Thus, the .79 Written Score reliability is actually slightly higher than what would have been originally projected from a 5 essay/1 PT examination.

SUMMARY AND CONCLUSIONS

The key findings from the evaluation of the first implementation of the modified 2-day CBX are:

- The modifications made to the bar examination had no differential impacts on the component scores, total test scores or passage rates of women versus men or white versus minority applicants. These findings substantiate the outcomes of the earlier simulations conducted prior to modification of the exam which predicted that passing rates within the respective gender and racial/ethnic groups would only be impacted by 1% to 2%, subsequently resulting in virtually no change in the relative passing rates of different genders or racial/ethnic groups.
- As also predicted in the earlier simulations, the overall reliability of the examination was not negatively impacted under the modified examination structure. Indeed, the test reliability actually improved, aided in part by recent increases in the reliability of the MBE and written sections.
- Under the new weighting scheme applied in the 2-day exam format, only .7% fewer applicants (n=198) passed the exam that would have otherwise passed under the previous weighting scheme. All of these applicants went into regrade and eventually passed, however.
- For the first time since 2008, CBX scores and passage rates reversed their downward trend in 2017. MBE scores rose by 9 points to 1432 in 2017, after reaching their lowest level in 2016 (1423). This reversal in the MBE was also observed nationwide, where scores increased by 14 points from 1403 to 1417. For California applicants, the scores on the Written Section, which is scaled to the MBE, and the Total Scale Scores also increased from previous years, and the July 2017 bar passage rate increased by 6% over July 2016.

- A separate examination of those repeating the bar for the first time during each of the two study timeframes (July 2015/July 2016 and July 2016/July 2017) was also conducted. This analysis revealed that the group of repeaters first taking the 3-day test in July 2016 and then repeating it in July 2017 under the new 2-day format exhibited statistically significant increases in their scores and passage rates when compared to those who took the 3-day version for both their first test cycle in July 2015 and again in July 2016. This is an intriguing finding since the two groups showed no statistically significant differences in their MBE scores from their first exams, suggesting that the two cohorts were of roughly equivalent ability.

The primary conclusion drawn from the first three of these findings is that, as predicted in the original simulations conducted to predict the outcomes of a shortened examination, implementation of a modified 2-day exam format for the CBX did not have any discriminatory impacts on applicants or negative effects on test reliability. These outcomes were achieved while lessening the testing burden on both applicants and the California State Bar organization.

Definitive conclusions regarding the reported increases in test scores and passing rates observed with the July 2017 implementation of the new 2-day format are less easily drawn. Since these improvements reverse a long downward trend in scores, and since test repeaters in 2017 performed better on the modified examination than the cohort of test repeaters re-administered the 3-day format for their second attempt, the obvious question to emerge is whether the modified examination was somehow “easier”. As previously discussed, the parallel improvement in MBE scores nationally in 2017, following a similar period of decline as observed in California, argues against this interpretation. Additional analyses also established that the change to the weighting scheme under the 2-day exam was not a factor.

Other possible contributing factors, including differences in applicants’ abilities and/or preparation, and the statistical phenomenon of regression toward the mean, have been discussed in the

report. The latter can be investigated with analyses of score data from future exams, but the investigation of other possibilities would require compilation and analyses of far more extensive data than was available at the time of this evaluation. Although we cannot provide a definitive explanation of the observed increases in scores, on the balance, we found nothing that pointed to the change in exam format as the primary cause of these improvements.

REFERENCES

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *The Standards for Educational and Psychological Testing* (2014).

Bolus R. (2017a). *Estimation of Passing Rates for a 2-Day California Bar Examination: An update of the 2011 report using February and July data from 2011 to 2016*. A report prepared for the California Committee of Bar Examiners.

Bolus R. (2017b). “2-Day Exam” *Simulation using the February 2017 examination*. A memorandum prepared for the Executive Director, California Bar Department of Admissions.

Buckendahl, C. (2013). *Key Factors to Consider when Engaging in a Development or Redevelopment Process for Examinations (California’s Two-Day Examination Proposal)*. A report prepared for the California Committee of Bar Examiners.

Holton, V. & Grunberg R. (2016). *Request that the Supreme Court Adopt Proposed Modifications to the California Bar Examination*. A report presented to the Supreme Court of California by the Office of General Counsel, The State Bar of California.

Klein, S. & Bolus, R. (2011). *The Estimated Effect on Examination Quality and Passing Rates of Different Ways of Modifying California’s Bar Examination*. A report prepared for the California Committee of Bar Examiners.

