



The State Bar *of California*

**OPEN SESSION
AGENDA ITEM
DECEMBER 2019
COMMITTEE OF BAR EXAMINERS ITEM O-201**

DATE: December 6, 2019

TO: Members, Committee of Bar Examiners

FROM: Lisa J. Cummins, Program Manager, Examinations

SUBJECT: Discussion of Report Prepared by the Committee's Psychometrician on the Impact of the Modified (2-Day) California Bar Examination

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 2017 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' March 2, 2018 report contained the following key findings:

- 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.

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

As a condition of its approval of the change in the exam format, the State Bar's Board of Trustees (BOT) requested that the impact be further evaluated after the 2-Day format had been in use for 2 years, i.e., *for four administrations* (i.e., two July and two February administrations). While the primary questions remained essentially the same as those posed in the March 2018 report, an additional focus would be on whether the initial findings would (a) remain consistent across multiple administrations of the examination and (b) would differ by July vs. February administrations where the groups of test-takers sitting for those exams historically had performed at different proficiency levels. Further, the BOT wanted to know, if the findings were found to be inconsistent, what was the nature of those differences and whether there were differential effects on examinee subgroups and their outcomes.

DISCUSSION

As mentioned, an initial study of the three core questions relating to the July 2017 CBX was conducted by Dr. Bolus and published in March 2018. The updated investigation replicated the analyses from that study, extending the sample to 25,717 examinees sitting for the first four administrations of the 2-Day format and 25,001 examinees who sat for the final four administrations of the 3-Day format (and served as the comparison group).

Dr. Bolus' August 31, 2019 "Report on the Impact of the Modification of the California Bar Examination from a Three-Day to a Two-Day Format" is attached for your review.

The following key findings that emerged from this evaluation include the following:

- The modifications made to the bar examination had no discernable impacts on test-takers component scores (i.e., essay, performance tests or MBE), total test scores or passage rates. The results were similar for both February and July administrations. These findings were further supported by the continuing relationship observed between California and the rest of the U.S. on MBE performance before the configuration change.
- Changes to the CBX format did not have any disparate impact on the performance of subgroups of interest, including females, minorities, students from non-ABA law schools or different tiers of California ABA law schools. Any changes that were observed accounted for less than .01% of the variation in scores or passing rates and could be attributable to chance alone.
- Upon repeating the examination, applicants who initially failed while taking the 3-Day format did no better than a similar group of applicants taking the 2-Day format.
- Based on modeling of the impact of changing the test section weighting, there was no discernable impact caused by reducing the weight of the written section from 65% to 50%. Only 140 (.5%) more applicants were passed than would have been estimated under the historic weighting scheme; with no difference at all on February administrations.
- Analyses estimated that 96% of the 25,000+ test takers would have achieved the same pass/fail status (based on the first phase of grading) if the weighting scheme had remained the same. On each of the four 2-Day administrations, an equal percentage (2%) were estimated to have switched outcome. The percentage of mis-matched applicants would be expected to be smaller after the regrading phase.
- Further analysis of the 974 applicants estimated to have changed pass/fail status using the revised weighting scheme, showed no systematic bias associated within any of the subgroups that were studied. That is to say, the representation of each subgroup (e.g., women) within 974 was statistically identical to their representation in the overall sample of 25,000+.
- While the reliability of the written portion of the examination dropped slightly from the 3-Day to 2-Day administration (average $r_{tt}=.79$ vs. $.76$), the overall test reliability actually increased (average $r_{tt}=.89$ vs. $.91$). The reason for the increase in reliability was a function of the additional weight given to the MBE under the 2-Day configuration.

These findings suggest that the conversion to a 2-Day format had negligible, if any, effect on the outcomes of applicants sitting for the CBX, while the change resulted in a reduction in the testing burden to both applicants and the California State Bar.

RECOMMENDATION

It is recommended that the Committee receive and file Dr. Bolus' August 31, 2019 report and authorize its publication.

PROPOSED MOTION

If the Committee agrees with the recommendation, the following motion should be made:

Move, that the Report on the Impact of the Modification of the California Bar Examination from a Three-Day to a Two-Day Format, prepared by Roger Bolus, Ph.D. and dated August 31, 2019, be received and filed and that the report be authorized for publication.

**A Report on the Impact of the Modification of the
California Bar Examination
From a Three-Day to a Two-Day Format**

Roger E. Bolus, Ph.D.

Research Solutions Group

August 31, 2019

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 period. The revised format would consist of five written essay questions and a single performance test given on one day (rather than six essay questions and two performance tests given over two days), and the Multistate Bar Examination (MBE) multiple-choice test given on a second day. In terms of scoring, equal weighting would be given to the shortened written section and MBE, a departure from the 3-Day format where the written section was given 65% weight. This recommendation was supported by the outcomes of statistical simulations which indicated that reducing the number of written and performance tests 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 Supreme Court's approval, the modified CBX was used for the first time for the July 2017 administration of the exam. As a condition of approval, the State Bar's Board of Trustees (BOT) requested an analysis of the results of this initial implementation of the 2-day format after it had been in use for two years (four administrations). The analyses were intended to address three core issues: (1) to what degree were the scores and subsequent bar passage rates impacted by the modified CBX? (2) to what degree were key demographic subgroups impacted? and (3) to what degree was the reliability of the CBX (and its components) impacted?

An initial study of these questions was conducted on the July 2017 CBX and published in March 2018. The current investigation replicated the analyses from that study, extending the sample to 25,717 examinees sitting for the first four administrations of the 2-Day format and 25,001 examinees who sat for the final four administrations of the 3-Day format (and served as the comparison group).

The following key findings that emerged from this evaluation include the following:

- The modifications made to the bar examination had no discernable impacts on test-takers component scores (i.e., essay, performance tests or MBE), total test scores or passage rates. The results were similar for both February and July administrations. These findings were further supported by the continuing relationship observed between California and the rest of the U.S. on MBE performance before the configuration change.

- Changes to the CBX format did not have any disparate impact on the performance of subgroups of interest, including females, minorities, students from non-ABA law schools or different tiers of California ABA law schools. Any changes that were observed accounted for less than .01% of the variation in scores or passing rates and could be attributable to chance alone.
- Upon repeating the examination, applicants who initially failed while taking the 3-Day format did no better than a similar group of applicants taking the 2-Day format.
- Based on modeling of the impact of changing the test section weighting, there was no discernable impact caused by reducing the weight of the written section from 65% to 50%. Only 140 (.5%) more applicants were passed than would have been estimated under the historic weighting scheme; with no difference at all on February administrations.
- Analyses estimated that 96% of the 25,000+ test takers would have achieved the same pass/fail status (based on the first phase of grading) if the weighting scheme had remained the same. On each of the four 2-Day administrations, an equal percentage (2%) were estimated to have switched outcome. The percentage of mis-matched applicants would be expected to be smaller after the regrading phase.
- Further analysis of the 974 applicants estimated to have changed pass/fail status using the revised weighting scheme, showed no systematic bias associated within any of the subgroups that were studied. That is to say, the representation of each subgroup (e.g., women) within 974 was statistically identical to their representation in the overall sample of 25,000+.
- While the reliability of the written portion of the examination dropped slightly from the 3-Day to 2-Day administration (average $r_{tt}=.79$ vs. $.76$), the overall test reliability actually increased (average $r_{tt}=.89$ vs. $.91$). The reason for the increase in reliability was a function of the additional weight given to the MBE under the 2-Day configuration.

These findings suggest that the conversion to a 2-Day format had negligible if any effect on the outcomes of applicants sitting for the CBX, while the change resulted in a reduction in the testing burden to both applicants and the California State Bar.

INTRODUCTION & BACKGROUND

Preceding July 2017, the California Bar Examination (CBX) was administered over a three-day period. Until that administration, California was one of only five states in which the bar examination exceeded two days in duration. The configuration¹ of the CBX until that time had consisted of six 1-hour essay questions, two 3-hour performance tasks and the 200-item Multistate Bar Examination (MBE) multiple choice test. Growing concern had been expressed by the public and the California Committee of Bar Examiners (CBE) that the length of the examination was causing undue burden on applicants to the Bar, and generating unnecessary administrative expenses. In response, the CBE commissioned research to evaluate (a) whether the CBX could be shortened without sacrificing its quality, (b) whether a shortened version would unduly impact any specific subgroup of test-takers, and (c) how might a shortened version be configured. The request was in line with the directives of *The Standards for Educational and Psychological Testing*, charging license testing agencies to continually consider methods for improving efficiencies in test development and administration, while maintaining psychometric quality.

In 2011, Klein & Bolus published the results of a study that evaluated multiple alternative configurations for the CBX. The study focused on alternatives for 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 examination configurations 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 these results, in 2015 the CBE recommended that the CBX be modified to include fewer essay and performance task questions, be administered over a two-day period, and adjustments to the weighting of the different test sections. Specifically, the following modifications were recommended:

1. MBE. Under the 3-Day format, the MBE was administered during the second day of testing, preceded and followed by two full days of written questions. Under the 2-Day format, the MBE would continue to be administered on the 2nd day of testing, but then the examination would conclude.

¹ We use the term format, configurations and administrations interchangeably

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

2. Essay and Performance Task. Under the 3-Day format, three essay questions were administered during a 3-hour morning session on the 1st and 3rd day of testing (a total of 6 questions) while two Performance Tasks were administered during the afternoons of the 1st and 3rd day of testing; each over a three-hour period. Under the 2-Day format, only 5 essay questions and one Performance Task would be administered in a single, slightly expanded day; three essays during a 3-hour morning session and two essays and a Performance Task during a 3.5-hour afternoon session. Thus, the Essay and Performance Task under the different formats would differ both in terms of number (six vs. five essays and two vs. one performance task) and the amount of time allocated (six vs. five hours for the essay tasks and six vs. 1.5 hours for the performance task³).
3. Total Scale Score Calculations. Under both testing formats, the raw written score (i.e., the weighted combination of essay and performance task) would be scaled to the MBE to arrive at a total scale written score, and then both summed to arrive at a total overall scale score. A major difference was that under 3-Day format, the MBE and written scores were given different weightings (65% for the written and 35% for the MBE). Under the 2-Day format, to adjust for the reduced number of and time dedicated to the written section, equal weights (i.e., 50%) would be assigned to each section, thereby giving more “importance” to the MBE in determining an applicant’s outcome.
4. Pass/Fail Decisions. Under the 3-Day format, an applicant would pass if they achieved a score of 1,440 (out of 2,000). It was proposed that no change be made to this rule, thus the standard for determining proficiency would not change.

Notice of the impending changes was given to the public in July of 2015. In March 2016 a formal request for adoption was sent to the California State Supreme Court (Holton & Grunberg, 2016). Following the Court’s approval, the modified CBX was used for the first time during the July 2017 administration.

Following the decision to change the examination, the Committee requested that results of the initial July 2017 implementation of the 2-day administration of the CBX be evaluated. In March 2018, a study was completed (Bolus, 2018). The primary questions investigated in that study were:

³ Under the 2-Day format, examinees could split up their 3.5-hour session in any way they choose. Recommendations are provided to students to use two hours for the essays and 1.5 hours for the performance task.

1. To what degree had the scores and subsequent bar passage rates on that exam been impacted by the modified CBX structure?
2. To what degree had the differences in scores and subsequent passage rates between key demographic subgroups been impacted?
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?

Based upon the results of that initial administration of the 2-Day format, the study found:

- The modifications made to the bar examination had no overall impact and no differential effects on the component scores (i.e., MBE or written section), total test scores or passage rates of women versus men or white versus minority applicants.
- Under the revised 50/50 weighting scheme, less than 1% fewer applicants passed the exam that would have otherwise passed under the previous 65/35 scheme.
- The overall reliability of the examination was not only not negatively impacted under the shortened examination structure, but actually improved, due in part to the increased reliability of the MBE.
- For the first time since 2008, CBX scores and passage rates reversed their downward trend, though the improvements that could not be directly linked to changes in the examination structure.
- Those applicants repeating the bar for the first time (i.e., 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 larger increases in their scores and passage rates than the group of repeaters those who took the 3-day version for both their first test cycle in July 2015 and again in July 2016.

The study concluded that, as predicted in the original simulations conducted to predict the outcomes of a shortened examination, the initial 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. The outcomes were achieved while reducing both the testing burden on applicants and the costs associated with administering the examination.

As a condition for adoption of the change, the CBE requested that the impact be further evaluated after the 2-Day format was in place *for four administrations* (i.e., two July and two February administrations). While the primary questions remained essentially the same as those posed in the March 2018 report, an additional focus would be on whether the initial findings would (a) remain consistent across multiple administrations of the examination and (b) would differ by July vs. February administrations where the groups of test-takers sitting for those exams historically had performed at different proficiency levels. Further the CBE wished to know that if the findings were found to be inconsistent, what was the nature of those differences and whether there were differential effects on examinee subgroups and their outcomes.

METHODOLOGY

Primary Research Questions.

This current study was designed to address the issues posed by the CBE by replicating the analyses conducted in the March 2018 report, while expanding the scope of analyses to include three additional examinations, the February 2018 and 2019, and the July 2018 administrations.

Specifically, the research sought answers to the following questions:

- Have scores and bar passage rates been impacted by the modified CBX structure? Specifically, did applicants find the new format any more or less difficult?
- If examinee performance was affected, did they originate with specific portions of the CBX?
- Were any observed differences in performance under the 2 vs. 3-day formats vary by different subgroups of test-takers (e.g., minority applicants), and if so, what were the magnitude of those differences?
- What was the impact of modifying the relative weights assigned to each of the test sections?
- Did the changes in the structure of the examination impact the psychometric quality of the examination?

Applicant Data.

To empirically answer these questions, data from eight administrations of the GBX were used in the analyses. The first set of data were taken from the initial four administrations under the revised 2-Day format (July 2017 and 2018, and February 2018 and 2019). A comparison group comprised of four administrations under the 3-Day format made up the second set of data (July 2015 and 2016, and February 2016 and 2017). Within each administration, only applicants with a full set of scored written questions and an MBE score were included. Both first time takers and repeaters were included.

Approximately 51,000 applicant records were extracted from the State Bar's historical files; 25,717 taking the 2-Day format and 25,001 taking the 3-Day format. For each applicant, raw scores, scale scores and final result (i.e., pass vs. fail) were included in the extract. In addition to examination results, the applicant's race/ethnicity, gender, law school and number of previous CBX's taken were included. All applicants were included in the population level analyses. For specific analyses regarding subgroups, applicants with missing demographic information (i.e., gender or race) were excluded. For gender, less than .1% of cases were missing a designation. For analyses focusing on race/ethnicity, four

groups were included (Asian, Hispanic, Black and White) making up 96% of all test-takers; the remaining 4% either were missing racial/ethnic information or had unclassifiable codes. Analyses focusing on applicant's law school were limited to those applicants graduating from one of the three categories of California Law Schools (ABA-Approved, Accredited, Unaccredited) or an out-of-state ABA-Approved School. Students from these schools made up about 82% of all test takers. The balance were predominantly lawyers taking a bar examination in the state. For analysis purposes, students from California ABA-Approved ABA law schools were further stratified, into one of three tiers, based upon the entering LSAT credentials of the schools they attended. Seven to eight law schools were included in each group.

Table 1 on the following page summarizes the counts of examinees included, stratified by the various applicant categorizations (e.g., demographics, repeater status) separately for the February and July administrations of the 3-Day and 2-Day administrations. Table 2 contains the same data expressed as percentages that the applicant counts within each category (e.g., 1st time test takers) made up of all applicants within the overall group (e.g. all February test-takers administered the 3-Day format).

Results from Tables 1 and 2 indicate that approximately the same number of examinees sat for the four administrations leading up to the 2-Day as sat for the first four administrations following the change. In terms of the composition of those examinees, the percentage of those who were taking the CBX for the first time decreased after the change (by 1.7% in February and 5.3% in July) as did the relative percentage of males (2.1% for both February and July administrations. We note that in terms of race/ethnicity, the percentage of whites also decreased (3.5% overall). With respect to the type of law school that the examinees attended, the relative percentage of students from out-of-state ABA dropped slightly (1.6% overall) while applicants from the top-tiered California ABA schools increased somewhat (1.4%) overall. While these changes in composition of test-takers are fairly small, they are for the most part, associated with groups that have historically performed higher on the CBX than the other groups within each of the categories (except for ABA School Tier). As a result, any aggregate differences in examinee outcomes under the 3 vs 2-Day formats needed to control for these various characteristics.

Table 1

Counts of Examinees Taking the

February and July California Bar Examination

Under the 3-Day and 2-Day Formats

By Examinee Category

<u>Category</u>	<u>Examination Length</u>					
	<u>3-Day</u>			<u>2-Day</u>		
	<u>February</u>	<u>July</u>	<u>All</u>	<u>February</u>	<u>July</u>	<u>All</u>
<u>Status</u>						
1st Timer	2,515	10,918	13,433	2,389	10,452	12,841
Repeater	6,602	4,966	11,568	6,839	6,037	12,876
All	9,117	15,884	25,001	9,228	16,489	25,717
<u>Sex</u>						
Missing	57	128	185	110	174	284
Female	4,635	7,940	12,575	4,858	8,558	13,416
Male	4,425	7,816	12,241	4,260	7,757	12,017
All	9,117	15,884	25,001	9,228	16,489	25,717
<u>Race</u>						
Asian	1,952	3,260	5,212	2,141	3,611	5,752
Hispanic	1,406	2,225	3,631	1,517	2,545	4,062
Black	731	1021	1,752	773	1194	1,967
White	4,399	8,190	12,589	4,150	7,904	12,054
Other	629	1188	1,817	647	1235	1,882
All	9,117	15,884	25,001	9,228	16,489	25,717
<u>Law School Type</u>						
CA ABA Approved	3,700	8,439	12,139	3,580	8,347	11,927
ABA Approved-Non-CA.	1,497	3,114	4,611	1,322	2,891	4,213
CA Accredited	1317	1,556	2,873	1331	1,640	2,971
CA Unaccredited	532	564	1,096	588	666	1,254
Other	2,071	2,211	4,282	2,407	2,945	5,352
All	9,117	15,884	25,001	9,228	16,489	25,717
<u>CA ABA Law School Tier</u>						
Low	1,772	2,639	4,411	1,656	2,597	4,253
Medium	1,119	2,669	3,788	1,101	2,531	3,632
High	809	3,131	3,940	823	3,219	4,042
Total	3,700	8,439	12,139	3,580	8,347	11,927

Table 2

Percentage of Examinees Taking the

February and July California Bar Examination

Under the 3-Day and 2-Day Formats

By Examinee Category

<u>Category</u>	<u>Examination Length</u>						<u>Difference</u>		
	<u>3-Day</u>			<u>2-Day</u>			<u>2-Day vs. 3-Day</u>		
	<u>February</u>	<u>July</u>	<u>All</u>	<u>February</u>	<u>July</u>	<u>All</u>	<u>February</u>	<u>July</u>	<u>All</u>
<u>Status</u>									
First-Time	27.6%	68.7%	53.7%	25.9%	63.4%	49.9%	-1.7%	-5.3%	-3.8%
Repeater	72.4%	31.3%	46.3%	74.1%	36.6%	50.1%	1.7%	5.3%	3.8%
<u>Sex</u>									
Female	51.2%	50.4%	50.7%	53.3%	52.5%	52.8%	2.1%	2.1%	2.1%
Male	48.8%	49.6%	49.3%	46.7%	47.5%	47.2%	-2.1%	-2.1%	-2.1%
<u>Race</u>									
Asian	21.4%	20.5%	20.8%	23.2%	21.9%	22.4%	1.8%	1.4%	1.5%
Hispanic	15.4%	14.0%	14.5%	16.4%	15.4%	15.8%	1.0%	1.4%	1.3%
Black	8.0%	6.4%	7.0%	8.4%	7.2%	7.6%	0.4%	0.8%	0.6%
White	48.3%	51.6%	50.4%	45.0%	47.9%	46.9%	-3.3%	-3.6%	-3.5%
Other	6.9%	7.5%	7.3%	7.0%	7.5%	7.3%	0.1%	0.0%	0.1%
<u>School Category</u>									
CA ABA Apprvd	52.5%	61.7%	58.6%	52.5%	61.6%	58.6%	0.0%	-0.1%	0.0%
Non CA ABA Apprvd	21.2%	22.8%	22.3%	19.4%	21.3%	20.7%	-1.9%	-1.4%	-1.6%
CA Accredited	18.7%	11.4%	13.9%	19.5%	12.1%	14.6%	0.8%	0.7%	0.7%
CA Unaccredited	7.6%	4.1%	5.3%	8.6%	4.9%	6.2%	1.1%	0.8%	0.9%
<u>CA ABA Tier</u>									
Low	47.9%	31.3%	36.3%	46.3%	31.1%	35.7%	-1.6%	-0.2%	-0.7%
Medium	30.2%	31.6%	31.2%	30.8%	30.3%	30.5%	0.5%	-1.3%	-0.8%
High	21.9%	37.1%	32.5%	23.0%	38.6%	33.9%	1.1%	1.5%	1.4%

The balance of this report presents data addressing each of the questions, interprets those results and attempts to draw conclusions from the findings.

RESULTS

This section of the report presents statistical data addressing each of the research questions presented above.

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

We first respond to this question by presenting data on the overall changes in performance of applicants sitting for the examination under the 2-Day vs 3-Day format. We present results on four performance CBX outcomes that may have been impacted by the change in format. In each analysis, we present mean (average) performance levels on the MBE (range of 0-2000 points), the average written score across questions and performance task (range of 40-100), the overall CBX Total Scale Score (range of 0-2000) as well as the percentage of applicants who passed the examination. We also calculate the differences in performance on each of the metrics by examination format.

To isolate differences and put them into some context, we apply a technique known as Analysis of Variance (ANOVA) modeling to assess the *amount of variation* in the outcomes (out of 100%) that can be uniquely attributed to the various factors that are being studied (e.g., test configurations, test month)⁴. Given the size of our sample (51,000+), we rely on this index rather than traditional tests of statistical significance that are overly sensitive to minute differences and fail to differentiate small statistical differences from those that are more meaningful and substantive.

Table 3 on the following page presents outcomes across the entire population of applicants sitting for the 2 and 3-Day administrations of the February and July examinations.⁵ Results from Table 1 reveal only minor decreases in overall performance on all outcome measures for both administrations, with the lone exception being a slight increase in written scores (.4 points on an average essay and .1 on an average performance task). Several pieces of evidence, however, suggest that these differences are non-significant and most likely due to random variation. First, the ANOVA analysis indicated that less than .1% of the overall variation in any of the five outcome measures could be attributed to the differences in examination configuration. Secondly, the magnitude of the differences that were observed

⁴ As a frame of reference for interpreting variation accounted for, in his 2018 report, Bolus found that Final Law School Grade Point Average (GPA) was the most significant predictor of CBX outcomes. It was found to account for fully, 44% of variance in CBX total scores.

⁵ To minimize the effect of random between administration variation, both the February and July results have been aggregated to include two administrations each.

in Table 3 did not deviate from normal historic, year-over-year changes. That is to say, on the ten July administrations preceding the modification, the average between-year difference on the MBE, essay, performance task, total score and percentage passing were 11, .9, .6, 11.1 and 2.8%, respectively. The differences observed in Table 3 (5, .4, .1, -5 and 1%, respectively) were actually smaller and well within the ranges of the historic distributions of year-over-year changes. Results were similar when compared to 10 historic February administrations.

Table 3
Average CBX Outcomes
Stratified by Month of Administration and
Type of Format*

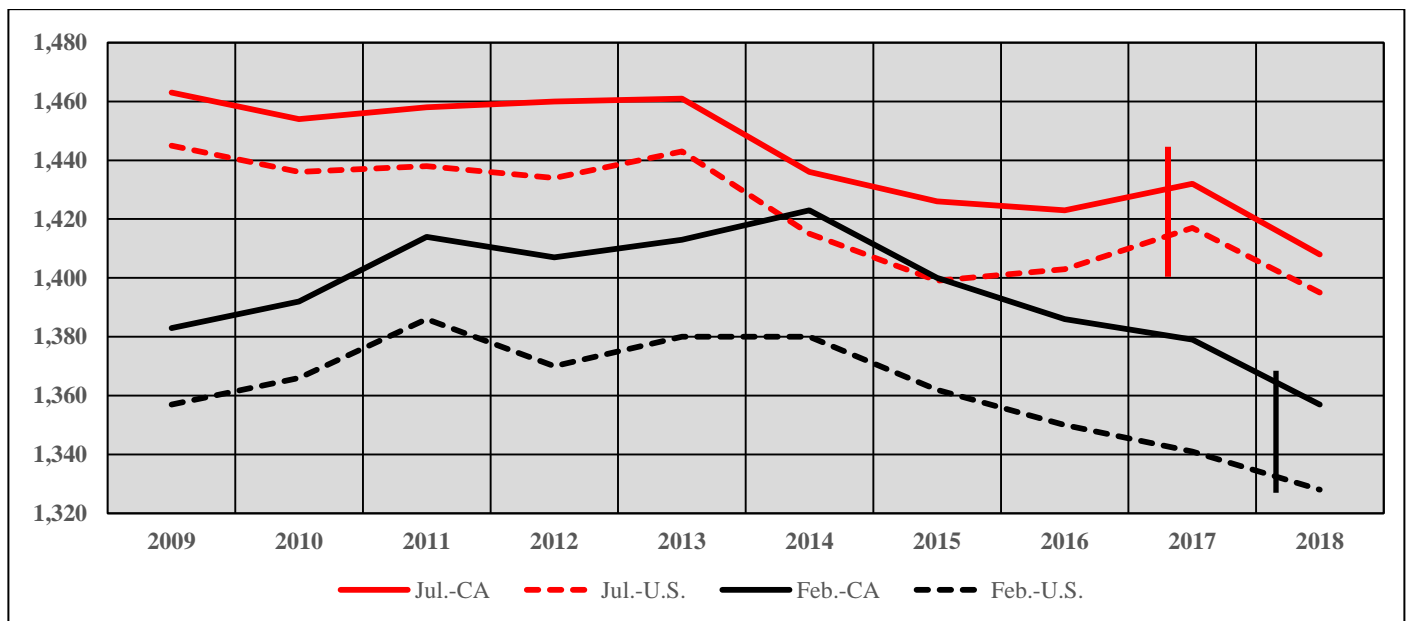
<u>Outcome</u>	<u>February</u>			<u>July</u>			<u>All</u>		
	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>
<u>MBE</u>	1384 (146)	1365 (153)	-19	1425 (162)	1420 (166)	-5	1410 (153)	1401 (164)	-9
<u>Written</u>									
<u>Essay</u>	59.9 (4.7)	59.8 (4.8)	-0.1	60.0 (5.2)	60.4 (5.6)	.4	59.9 (5.1)	60.2 (5.3)	.3
<u>PT</u>	61.3 (5.7)	59.4 (6.8)	-1.9	61.2 (6.3)	61.3 (8.1)	.1	61.2 (6.1)	61.3 (7.7)	.1
<u>Total</u>	1382 (132)	1365 (137)	-17	1423 (150)	1418 (153)	-5	1408 (145)	1399 (150)	-9
<u>% Pass</u>	35%	30%	-5%	45%	46%	1%	42%	40%	-2%

* Entries in parentheses are standard deviations

A final piece of supporting evidence can be found in the comparison of scores to national secular trends. Figure 1 illustrates the mean performance on the MBE by both California examinees and all U.S. test takers for both the July and February administrations from 2009 through 2018. Preceding the

configuration change (July 2016 and February 2017), California's performance, while higher than the national levels, tracked almost perfectly with the national levels⁶

Figure 1
Average Performance on the MBE
All U.S. vs. California
2009 through 2018*



* U.S. data not available for February 2019

The implementation of the 2-Day format apparently had no impact on the U.S./CA relationship. The average post-implementation MBE performance in California was within a few points of exactly what would have been predicted had the change not taken place.

Thus, the preponderance of data suggests that the change in the CBX structure from a 3-Day to 2-Day configuration, in and by itself, did not affect the overall performance on any of the examinee outcomes.

Research Question 2. Did the 2-Day format impact some groups of applicants differently than others?

⁶ The correlation between average U.S. performance on the MBE and that of California over the years is .98 and .96 for the July and February administrations respectively.

The above analyses indicated that the move to a 2-Day format did not have an overall impact on the 25,000+ test-takers taking the examination after the implementation. However, the question remains as to whether specific subgroups may have been differentially impacted by the change. Subgroups of interest include females, ethnic minorities, students graduating from non-ABA approved schools, or even students from lower tiered ABA-approved schools within California. *The question to be answered here is not whether the groups perform better or worse than other groups, but rather, did the gap in performance widen or narrow with onset of the revised format.* And if so, could this effect be attributed to changes (e.g., configuration or testing time allocations) on one portion of the examination or both? And finally, even if actual scores were impacted, was there a net effect on the actual passing rates of those subgroups?

Method. To answer these questions, we conducted a series of ANOVA analyses. For each performance outcome (e.g., average performance task score), we started with a baseline model that first determined the percentage of variation in scores that could be accounted for uniquely by each of two factors; (a) the format under which the applicant sat for the exam (i.e., 2-Day vs. 3-Day configurations⁷) and (b) the subgroup category of interest (e.g., gender). We then estimated the amount of variation in the outcomes that could be accounted for by these two factors. A second model was then developed which included the two factors, but add the *interaction* of the two. The interaction was included to evaluate whether the size of any differences that existed between different groups (e.g., females vs. males) under the 3-Day format was equivalent to the difference under the 2-Day format. A significant interaction would suggest that those differences did increase (or decrease) and would be evidenced by an increase in the amount of variation accounted for in the expanded model over the baseline model. No, or minimal increase in explained variation would suggest that the size of differences in performances between the groups did not change under the 2-Day format and thus, there would be no evidence of differential impact. This modeling was conducted separately for each outcome metric (MBE, average essay and performance test score, total scale score and passing rates) for each student demographic and category mentioned above.

Tables 4 through 9 present the CBX outcomes before and after the configuration changes. Each table stratifies the results for each of the outcomes by separate subgroups (aggregated across February/July administrations). In addition to the average performance under each CBX configuration, the tables present the calculation of mean differences between the two.

⁷ Even though our initial analyses showed no effect, we included the factor for the sake of completeness)

Table 4**Average CBX Outcomes****Stratified by Type of CBX Format and Gender**

<u>Outcome</u>	<u>Female</u>			<u>Male</u>		
	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>
<u>MBE</u>	1,392	1,385	-7	1,427	1,418	-9
<u>Written</u>						
<u>Essay</u>	60.2	60.5	0.3	59.7	59.9	0.2
<u>PT</u>	61.6	61.0	-0.6	60.8	60.2	-0.6
<u>Total</u>	1,408	1,395	-13	1,407	1,402	-5
<u>% Pass</u>	42%	39%	-3%	41%	41%	0%

Table 5**Average CBX Outcomes****Stratified by Type of CBX Format and Racial/Ethnic Group**

<u>Outcome</u>	<u>Asian</u>			<u>Hispanic</u>			<u>Black</u>			<u>White</u>		
	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>
<u>MBE</u>	1,380	1,372	-8	1,381	1,378	-3	1,357	1,348	-9	1,440	1,433	-7
<u>Written</u>												
<u>Essay</u>	59.2	59.1	-0.1	59.4	59.9	0.5	58.0	58.4	0.4	60.7	61.2	0.5
<u>PT</u>	60.9	59.7	-1.2	60.4	60.0	-0.4	58.8	58.3	-0.5	61.9	61.6	-0.3
<u>Total</u>	1,385	1,367	-18	1,383	1,381	-2	1,343	1,342	-1	1,435	1,430	-5
<u>% Pass</u>	37%	34%	-3%	35%	34%	-1%	22%	23%	1%	49%	48%	-1%

Table 6

Average CBX Outcomes

Stratified by Type of CBX Format and Examinee Law School

<u>Outcome</u>	<u>CA ABA Approved</u>			<u>Non-CA ABA</u>			<u>CA Accredited</u>			<u>CA Unaccrcdited</u>		
	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>
<u>MBE</u>	1,444	1,436	-8	1,429	1,426	-3	1,327	1,318	-9	1,339	1,337	-2
<u>Written</u>												
<u>Essay</u>	61.4	61.8	0.4	59.7	60.2	0.5	57.9	58.5	0.6	56.9	57.7	0.8
<u>PT</u>	62.6	62.4	-0.2	61.9	61.7	-0.2	58.3	57.7	-0.6	57.3	56.8	-0.5
<u>Total</u>	1,452	1,443	-9	1,418	1,416	-2	1,326	1,324	-2	1,309	1,321	12
<u>% Pass</u>	54%	51%	-3%	44%	44%	0%	15%	16%	1%	12%	14%	2%

Table 7

Average CBX Outcomes

Stratified by Type of CBX Format and Examinee Law School Tier*

<u>Outcome</u>	<u>Bottom Tier</u>			<u>Middle Tier</u>			<u>Top Tier</u>		
	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>
<u>MBE</u>	1,391	1,378	-13	1,449	1,438	-11	1,497	1,497	0
<u>Written</u>									
<u>Essay</u>	60.4	60.6	0.2	61.3	61.8	0.5	62.7	63.2	0.5
<u>PT</u>	60.9	60.4	-0.5	62.8	62.6	-0.2	64.3	64.4	0.1
<u>Total</u>	1,403	1,390	-13	1,454	1,445	-9	1,503	1,498	-5
<u>% Pass</u>	37%	34%	-3%	56%	52%	-4%	70%	68%	-2%

* Limited to California ABA schools

Table 8

Average CBX Outcomes

Stratified by Type of CBX Format and Number of Exams Taken

<u>Outcome</u>	<u>1st Time Taker</u>			<u>Repeater</u>		
	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>	<u>3-Day</u>	<u>2-Day</u>	<u>Diff.</u>
<u>MBE</u>	1,448	1,446	-2	1,365	1,355	-10
<u>Written</u>						
<u>Essay</u>	61.0	61.5	0.5	58.6	58.9	0.3
<u>PT</u>	62.6	62.5	-0.1	59.7	58.6	-1.1
<u>Total</u>	1,451	1,446	-5	1,357	1,351	-6
<u>% Pass</u>	56%	56%	0%	26%	24%	-2%

Results. While the data in the above tables indicate that changes in levels of performance do differ somewhat by applicant categories, analysis of those results suggest several common findings. First, and foremost, in none of the tables did we find *a significant differential CBX performance impact in any of the subgroups that could be directly attributable to the introduction of the 2-Day format*. In every model, for every outcome, the amount of variation added by any of the interactions of examinee subgroup by CBX configuration (to the baseline model) never exceeded .01%. In most instances the effects were not statistically indistinguishable from zero. For example, in the racial/ethnic baseline models examining the MBE outcome, the two factors (race/ethnicity and configuration) accounted for 4.103% of the total variation in MBE scores. The inclusion of the interaction of the two factors resulted in a model that accounted for 4.104% of the variation; an increase of only .001%! This finding was illustrated in the marked similarities in MBE mean scores differences under the 3-Day vs. 2-Day configurations within each racial/ethnic group (i.e. 8 points for Asians, 3 points for Hispanics, 9 points for Blacks and 7 points for White/Anglo). Further, these were all well within the range of normal, historic year-over-year changes in MBE performance under the 3-day format.

A second commonality among the various subgroup findings was that none of the baseline models independently accounted for much of the variation on any of the performance metrics among the 51,000+ examinees. As might be expected from historical trends, models that included category of law school and repeater status accounted for the greatest amount of variance (12% and 10%, respectively). Further, when all categories were included in a single model (including administration month), the

baseline model accounted for about 16%, while inclusion of *all* interactions added less than .3%, further underscoring the lack of impact of the change in CBX configuration.

Performance of Repeating Applicants. The above analyses have documented that the change in CBX configuration had no overall impact on examinee performance nor was there any differential impact that could be attributed to relevant subgroups of test-takers. One drawback of these analyses however is that other than repeaters, the applicants taking the exams differed between periods. While we attempted to account for possible differences in examinee ability, we reasoned that additional insights into the impact of the change in format of the CBX could be gained by looking more closely into the performance of examinees *who repeated the CBX*. To do this analysis, we created two groups of repeaters. The first group (N=1,797) was comprised of 1st time takers from July 2015 who failed the exam and then repeated in one of the 3 subsequent administrations 3-Day administrations. A second groups (N=1,535) was comprised of 1st time takers from July 2017 who failed the exam and then repeated in one of the 3 subsequent administrations 2-Day administrations⁸. For each group, we calculated the improvement in each of the CBX performance outcomes (i.e., the difference between 1st time scores and those achieved on their next attempt) to determine whether they were significantly larger under one format vs. the other. Similar to the previous models, we evaluated the percentage of variation in the differences that could be attributed to CBX format

A summary of the findings is presented in Table 9. The data in Table 9 represent the average repeater's performance on each metric on their first and subsequent attempt, along with the difference between the two. An average positive difference indicates improved performance. We also present the percentage of repeaters in each group that passed on their next attempt. Results in the table indicate that, as a group, the failing applicants taking the CBX under the 3-Day format scored slightly higher on the MBE, essay, and overall than those taking the CBX under the 2-Day format. Due to the wide variation in difference scores over time on each metric however, the impact of CBX configuration change account for *less than .6% of the overall variance* in outcome change. We did observe a slightly larger difference in the subsequent passing rates of those repeating under the 3-Day configuration (44% vs. 30%; a net difference of 14%). This finding is most likely attributable due to the cumulative improvement in both

⁸ An attempt was made to create a third group of 1st time takers of a 3-day format who repeated the CBX for the first time under a 2-Day format. Because the final 3-Day format was a February administration, there was an insufficient number of like examinees in this group to include in any analyses.

MBE and written performance. Yet, still only 2% of that difference in the respective total scale score and passing rate increase could be directly attributable to the examination structure.

Table 9
Failing July 2015 and 2017 1st Time Examinees:
Average Performance of On Their Initial and 2nd Attempt
Stratified by the Format of CBX

	<u>3-Day Only</u> (N=1,797)	<u>2-Day Only</u> (N=1,535)
<u>MBE</u>		
Initial Attempt	1,332	1,312
2nd Attempt	1,399	1,362
Difference	67	50
<u>Essay</u>		
Initial Attempt	57.4	57.8
2nd Attempt	59.7	59.8
Difference	2.3	2.0
<u>PT</u>		
Initial Attempt	59.5	58.8
2nd Attempt	62.1	60.2
Difference	2.6	1.4
<u>Total</u>		
Initial Attempt	1,334	1,320
2nd Attempt	1,406	1,367
Difference	72	47
Passing*	44%	30%

*On the 2nd attempt

Further analysis of repeater performance on future examinations may be warranted as bar passage rates begin to stabilize.

Research Question 3. What was the unique impact of modifying the relative weights assigned to each of the test sections?

In order to maintain the historic high levels of overall examination reliability with a shortened written section, it was necessary to increase the weight of the MBE in the 2-day format (.50) from what it was in the 3-day format (.35). The reasoning for this was that a shortened written section (five essays and one performance task) would, by definition, be less reliable than the longer version (six essays and two performance tasks). Critics of this change suggested that the revised format might favor test takers that perform better on multiple choice tests and net out to an overall increase in passage rates, and potentially lead to biased results. Others, who felt that the written section of the examination was the more valid of the two sections, felt that lowering the weighting on that section (from 65% to 50%) would undermine its perceived importance, and would lead to fewer woman passing (since they historically outperformed men on this section). The simulations and modeling that were conducted in the research leading up to the change predicted that there would be no effect on passage rates attributable to the weighting change, nor would it adversely impact any groups in any material manner.

To investigate this question, in the March 2018 report we conducted an analysis on the July 2017 *data that reversed the section weighting to their historic values* (i.e. .65 on the shortened written section at .65 and .35 for the MBE). Using the original written score scaling parameters, we re-calculated a total score and then compared scores and pass/fail rates under both scoring methods. The analysis revealed that less than 1% fewer applicants passed under the 50/50 weighting scheme than under the previous 65/35 weighting. While the finding did lend evidence to the lack of impact of the change in weighting schema, it was limited in that (a) it was conducted on a single examination without evidence that the results were replicable and (b) did not systematically explore if some subgroup was over-represented among the small sample of test-takers who were “reclassified (i.e., changed status from a pass to a fail, or vice-versa) as a result of score recalculation.

We replicated that analyses here, using the same methods, but expanding them to four 2-Day administrations to determine if the results would remain the same. As in the March 2018 report, our calculations are based on Phase 1 scores (though we do comment later on post-regrade outcomes). Table 10 presents the statistics on the number and percentage passing under the old weighting scheme (65/35) and the revised scheme (50/50). Data on both the total number passing and the percentage of total applicants passing are presented. The data in Table 1 show that 140 fewer applicants out of 25,700+ (i.e., about 5 out every 1,000 test takers) would have been estimated to pass over the four

administrations if the historic weights were applied. This, represents less than .5% difference. The entire difference was observed on the July administrations where there is generally more variation in scores than on the February administrations. We would expect that after subsequent grading phases, the results most likely have been even closer, as some of those who were close to the 1,440-passing standard (and failed in our simulation), most likely would have passed upon regrade.

Table 10
Actual and Estimated Passing Rates
Under the Current and Historic
Section Weighting Schemes
on the CBX 2-Day Administrations

<u>Exam</u>	<u>Passing-50/50</u>		<u>Passing-65/35</u>		<u>Difference</u>	
	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>
Jul 2017	4,105	48%	4,037	47%	68	0.8%
Jul 2018	3,178	40%	3,106	39%	72	0.9%
Average	3,642	44%	3,572	43%	70	0.8%
Feb 2018	1,219	26%	1,223	26%	-4	- 0.1%
Feb 2019	1,369	30%	1,365	30%	4	0.1%
Average	1,294	28%	1,294	28%	0	0.0%
All	9,871	38%	9,731	38%	140	0.5%

While it is clear that the overall passing rate would most likely not have been impacted, the question remains as to how many *individual* examinees might have been affected. We define an effected applicant as one who passed under one weighting system, but failed under the other, and vice-versa. To address this question, we compared an examinees' outcome under each system, and then calculated the percentage that would be estimated to have the same (i.e., pass under both, or fail under both) or a different (e.g., pass under the current system but fail under the historic) outcome. We present both the counts and percentages in Table 11, where we examine both the stability of the findings over both February and July administrations.

Table 11**A Comparison of Actual and Expected Examinee Outcomes****Under the Current and Historic****Section Weighting Schemes****on the CBX 2-Day Administrations**

<u>Exam</u>	<u>Same Outcome</u>					<u>Different Outcome</u>				
	<u>50/50 Pass- 65/35 Pass</u>		<u>50/50 Fail- 65/35 Fail</u>		<u>Total</u>	<u>50/50 Pass- 65/35 Fail</u>		<u>50/50 Fail- 65/35 Pass</u>		<u>Total</u>
	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>	<u>%</u>	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>	<u>%</u>
Jul 2017	3,907	46%	4,311	50%	96%	198	2%	130	2%	4%
Jul 2018	3,004	38%	4,663	59%	97%	174	2%	102	1%	3%
Average	3,456	42%	4,487	55%	96%	186	2%	116	1%	4%
Feb 2018	1,129	24%	3,341	72%	96%	90	2%	94	2%	4%
Feb 2019	1,274	28%	3,114	68%	96%	95	2%	91	2%	4%
Average	1,202	26%	3,228	70%	96%	93	2%	93	2%	4%
All	9,314	36%	15,429	60%	96%	557	2%	417	2%	4%

The results in Table 11 show a remarkable consistency. Fully, 96% of CBX examinees would have been predicted to have the identical outcome to their actual outcome if the weighting schema had not been changed. The results were identical for both of the July and February administrations. Further, of the 4% of examinees who were predicted to have a change in status, there was an equal split of those who passed but would have been predicted to fail (2%) and vice-versa (2%). These findings were not unexpected given the overall similarities in expected vs. actual passing rates, and suggest that those who did switch had actual and predicted scores that hovered around the 1,440-passing standard.

Finally, we investigated whether the 974 applicants who would have been predicted to change passing/failing status, were disproportionately members of some specific subgroup. That is to say, we sought to determine whether the weighting schema resulted in a biased outcome? To answer this question, for each subgroup of interest, we examined the outcome of these 974 examinees to assess whether the proportion of applicants who fell into the group (e.g., Blacks), were equivalent to their proportion in the entire population. A disproportionate rate would point to evidence of some bias. We

subjected each of the comparisons to a statistical Chi Square test of significance. A statistically significant Chi Square test would suggest that there *might* be some bias, and would warrant a closer look at the proportions to determine if the difference was substantively meaningful.

The results showed that for the two groups of primary interest, race/ethnicity and gender, neither statistical test was remotely significant ($\chi^2=3.2$; $p=.51$ for race/ethnicity and $\chi^2=.14$ for gender; $p=.71$) suggesting no evidence of disparate impact. When we looked at the same data by type of law school attended, we did see a small effect ($\chi^2=7.2$; $p=.016$), characterized by 1.5% more students from ABA schools being classified differently than those from other school types. However, this would not be unexpected because of the significantly larger number of students from California Accredited and Unaccredited schools scoring much lower (i.e., well below the passing standard) than those from ABA schools. This interpretation was substantiated when we found no disparate impact among the students from the three tiers of ABA law schools ($\chi^2=1.9$; $p=.37$).

Thus, consistent with the expected results from the earlier research leading up to the adoption of the 2-Day CBX configuration, and the findings from the assessment of the initial 2-Day implementation we can safely conclude from these analyses that the change in weighting schema has most likely not impacted the passing rates of the CBX. And, consistent with the early projections, the change in weighting schema does not appear to have affected any one particular subgroup of applicants at the expense of another.

Research Question 4. To what degree was the reliability of the CBX (and its components) impacted by shortening the length of examination and modifying the weighting of the respective sections?

The reliability⁹ of examination scores is impacted primarily by three 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 the 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 quite high. 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., potential differences in grader standards and practices, and inability to pilot test the questions before they are administered.

In terms of reliability, under the 3-Day configuration it was psychometrically warranted to apply a higher weighting to the written section than the MBE because under that configuration, there were 8 questions (six essays and two performance tasks). In arriving at the decision to shorten the CBX to two days, it was necessary to reduce the number of written questions in order to fit the time constraints. The only way that a high level of reliability could be maintained for the overall 2-day test format was to lower the weighting of the shortened written section and increase the weighting of the MBE. The prior simulation modeling undertaken while studying the potential impact of a two-day examination revealed that equal weighting of each section could not only maintain overall test reliability but possibly improve it.

In the 2018 report, it was shown that despite its shortened length, the *written section* reliability on the July 2017 CBX did not change significantly ($r_{tt} = .79$ vs $.83$ and $.82$ in the preceding 3-Day administrations), and because of the high level of reliability of the *MBE* ($r_{tt}=.93$), the actual *overall test reliability* increased ($r_{tt}=.92$ vs $.91$ and $.90$ on the preceding examinations). The recommendation in that report was to replicate these analyses to determine whether (a) the results were replicable with subsequent administrations, and (b) the results found on the July administrations, where there is

⁹ Reliability is a measure of the expected stability or reproducibility of a score. The reliability is represented by the symbol r_{tt} and ranges from .00 to 1.00; the closer the reliability index is to 1.00, the greater the confidence that the same score would be earned on a subsequent testing. For high stakes testing such as the CBX, it is desirable to have a r_{tt} that exceeds .85 to .90).

generally greater score variability were similar on February administrations, where the performance is generally lower and scores tend to cluster more (i.e., the standard deviations are smaller).

Towards this end, we calculated the reliabilities for each of the eight examinations under study. For each administration, we obtained the reliability of the MBE from the National Conference of Bar Examiners. We estimated the reliability of the written section based on both the essay and performance task scores. Finally, we estimated the reliability of the composite total score based on (a) the correlation of the respective sections, (b) the reliabilities of each of the sections, and (c) the weights assigned to each of those sections. Table 12 presents the results of those calculations, along with averages of the respective July and February administrations, and then the averages across the four CBX's administered under each of the two configurations.

Table 12
Section-Specific and Total Score Reliabilities

<u>Examination</u>	<u>MBE/Written Correlation</u>	<u>Reliability</u>		
		<u>MBE</u>	<u>Written</u>	<u>Total</u>
<u>3-Day</u>				
Jul '15	.70	.92	.83	.90
Jul '16	.73	.93	.82	.91
Ave.	.72	.93	.82	.91
Feb '16	.61	.90	.76	.86
Feb '17	.61	.92	.77	.87
Ave.	.61	.91	.77	.87
All 3-Day Exams	.66	.92	.79	.89
<u>2-Day</u>				
Jul '17	.72	.93	.79	.92
Jul '18	.71	.92	.79	.92
Ave.	.72	.93	.79	.92
Feb '18	.64	.92	.74	.90
Feb '19	.61	.92	.74	.89
Ave.	.63	.92	.74	.89
All 2-Day Exams	.67	.92	.76	.91

The findings in Table 12 confirm the findings from the 2018 report and reveal that for both February and July administrations of the CBX, *not only did the reliability of a shortened version of the CBX not suffer, it actually increased.* The overall total score reliability for July and February administrations under the 2-Day format averaged .92 and .89 respectively, compared to .91 and .87

under the 3-Day administrations. This occurred despite the fact that the reliability of the written section not unexpectedly decreased on average by .03 r_{tt} points. Increasing the weighting of the MBE (which averaged .92 reliability during each of the two time periods) had the affect identified during the early exploration of alternative formats for the CBX. Thus, the changes to the CBX realized the intended effect of reducing both the costs of administration and burden of an extra day of testing, without sacrificing the psychometric quality of the examination, adversely impacting passing rates, or causing disparate outcomes in any demographic groups.

SUMMARY AND CONCLUSIONS

Previous research on the implementation of a 2-Day administration of the CBX suggested that none of the various changes to the format or scoring had a deleterious impact on the quality of the examination or on its outcomes. Limited to study of a single administration, it was unclear whether those findings could be replicated in a more extensive study. The current investigation extended the research to include not only the first two years of experience and over 25,000 2-Day test takers, but added both a broader baseline period and added February administrations (which often vary from studies of July administrations due to difference in the characteristic of those test-takers).

The key findings from this expanded evaluation of the implementation of the modified 2-day structure of CBX were:

- The modifications made to the bar examination had no discernable impacts on test-takers component scores (i.e., essay, performance tasks or MBE), total test scores or passage rates. The results were similar for both February and July administrations. These finding were further supported by the similarity in the relationship between California and the rest of the U.S. on the MBE performance established before the configuration change.
- The modifications did not have any disparate impact on the performance of subgroups of interest, including females, minorities, students from non-ABA law schools or different tiers of California ABA law schools. Any changes that were observed accounted for less than .01% of the variation in scores/passing rates and could be attributable to chance alone.
- Upon repeating the examination, applicants who initially failed while taking the 3-Day format did no better than a similar group of applicants taking the 2-Day format. Investigation of subsequent failing (and re-taking) applicants is recommended.
- Upon modeling the impact of changing the test section weightings, we found no effect of reducing the weight of the written section from 65% to 50%. Only 140 (.5%) more applicants were passed than would have been estimated under the historic weighting scheme; with no difference at all on February administrations.
- Analyses also demonstrated that 96% of the 25,000+ test takers would have achieved the same pass/fail status (based on the first phase of grading) if the weighting scheme had remained the

same. On each of the four 2-Day administrations, an equal percentage (2%) were estimated to have switched outcome. We would expect the percentage of mis-matched applicants to be smaller upon regrading.

- Further analysis of the 974 applicants estimated to have changed pass/fail status on the revised weighting scheme, showed no systematic bias associated with any of the subgroups that were studied. That is to say, their representation within the sample of 974 was statistically identical to their representation in the overall sample of 25,000+.
- While the reliability of the written portion of the examination dropped slightly from the 3-Day to 2-Day administration (average $r_{tt}=.79$ vs. $.76$), the overall test reliability actually increased (average $r_{tt}=.89$ vs. $.91$). The average reliability of the MBE remained the same during that period ($r_{tt}=.92$) as did the correlation between sections ($r=.66/.67$). The reason for the increase in reliability was a function of the additional weight given to the MBE under the 2-Day configuration.

Taken in total, these findings suggest that conversion to a 2-Day CBX format, consisting of five written essays, a single, shortened performance task, and the MBE had negligible if any effect on the applicants sitting for the exam. Further, this favorable outcome was achieved while lessening the testing burden on both applicants and the California State Bar.

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.
- Bolus R, (2018a). *Analysis of the First Two-Day Administration of the California Bar Examination*. A report prepared for the California Committee of Bar Examiners.
- Bolus R. (2018b). *Performance Changes on the California Bar Examination: Part 2-New insights from a collaborative study with California Law Schools*. A report prepared for the State Bar of California.
- 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.