2017

Recent Performance Changes on the California Bar Examination (CBE):

Insights From CBE Electronic Databases

Results from the 2008, 2012 and 2016 administrations of the CBE are analyzed to determine the impact of the examination and examinee characteristics on the decrease in scores and passage rates

DR-01-17

Roger Bolus, Ph.D.
Research Solutions Group
2/19/2017

TABLE OF CONTENTS

SECTION	PAGE
EXECUTIVE SUMMARY …….	ii
I. BACKGROUND ……..	1
II. PURPOSE OF THE STUDY ……..	3
III. METHODS ……..	4
A. Study Data	4
B. Research Questions	6
IV. RESULTS ……..	9
1. Change in composition of test-takers	9
2. Change in scores and passing rates	12
3. Change in score distributions	19
4. Change in likelihood of passing after two years	24
5. Change in psychometric properties	27
6. Estimated change in passing rates under alternative standards	28
V. SUMMARY AND CONCLUSIONS …….	31

EXECUTIVE SUMMARY

The recent sharp drop in the percentage of applicants passing the Bar Examination in multiple states, including California, has generated considerable public debate regarding possible causes. The California Committee of Bar Examiners requested that the Research Solutions Group (RSG) conduct analysis of existing California Bar Examination (CBE) databases to: 1) establish a statistical baseline to profile the changes in passing rates that have occurred in California; and 2) determine if any insight could be provided from these databases into the factors that might have contributed to the decline in scores. On the basis of the data available, six research questions were posed to guide the analyses.

Data from the 2008, 2012 and 2016 examinations were analyzed. Over this 9 year period the following changes occurred:

- The number of test takers declined by 6% including an 11% decline in the number of July test takers and a 4% increase in February examinees.

- The mix of examinees shifted, with traditionally higher performing groups making up proportionately less of the total test takers over time.

- For the July exams, overall average Total Scale Scores (TSS) and bar passage rates dropped between 2008 and 2016: The average TSS declined 66 points (1481 to 1415) points and the percentage passing was 18% lower (62% to 44%) in 2016 than in 2008. Less pronounced decreases also occurred in the February exams between the two years.

The magnitude of the changes was not equal for all subgroups within applicant populations. The passing rate for applicants from CA ABA schools with higher median LSAT scores dropped 11% between 2008 and 2012 as compared to an almost 30% decrease for applicants from lower LSAT schools. The drop in passage rates in the various racial/ethnic groups varied by only 5% however. Additionally, the drop in scores on the Written and MBE sections were roughly equivalent within the various groups, suggesting that neither section disproportionately contributed to the change.

Results from an estimation model indicated that all things being held equal, roughly 20% of the change in July CBE scores and 17% of the change in bar passage rates could be attributed to the change in the mix of test takers between 2008 and 2016. Analyses also revealed a highly disproportionate number of test takers scored at the very lowest levels of the score distribution in 2016 relative to 2008 (21% vs 10%). A comparison of the composition of test takers scoring in this bottom portion of the distribution also revealed a disproportionate change across selected subgroups.

An analysis of “two-year” pass rates indicated that gaps in performance between 2008 and 2016 narrowed considerably, while an analysis of the reliability of the CBE actually showed very slight improvements on both the individual sections and overall scores.

Finally, analyses were conducted to evaluate the impact of alternative passing standards, or “cut points”, upon the decline in passage rates. If the modal U.S. standard of 135 were applied instead of California’s existing standard of 144, it is estimated that 22%

more applicants would have passed the July 2016 CBE. The size of the decrease between 2008 and 2016 would have shrunk by 3%. Finally, if California were to use a standard of 133 (the passing score applied in New York state), the decrease in passing rates between 2008 and 2016 was estimated to be identical (9%) to New York's for similarly situated applicants. Since New York adopted the Uniform Bar Examination in 2016 this finding suggests that use of the UBE format in California would probably have had little to no effect on the decrease in bar passage. Further, the change in passing rates for 1st time students from California ABA schools between 2008 and 2016 were similar to other states with large applicant pools.

These results suggest that there are most likely other factors beyond those examined in these analyses which are affecting the CBE passage rate. Institutional factors such as changes in curriculum and/or variation in student characteristics such as motivation, preparation and/or latent legal ability and law school performance may be operating. In the absence of additional data, however, we cannot asses the impacts of such variables. The nature, size and directionality of these decreases require additional data.

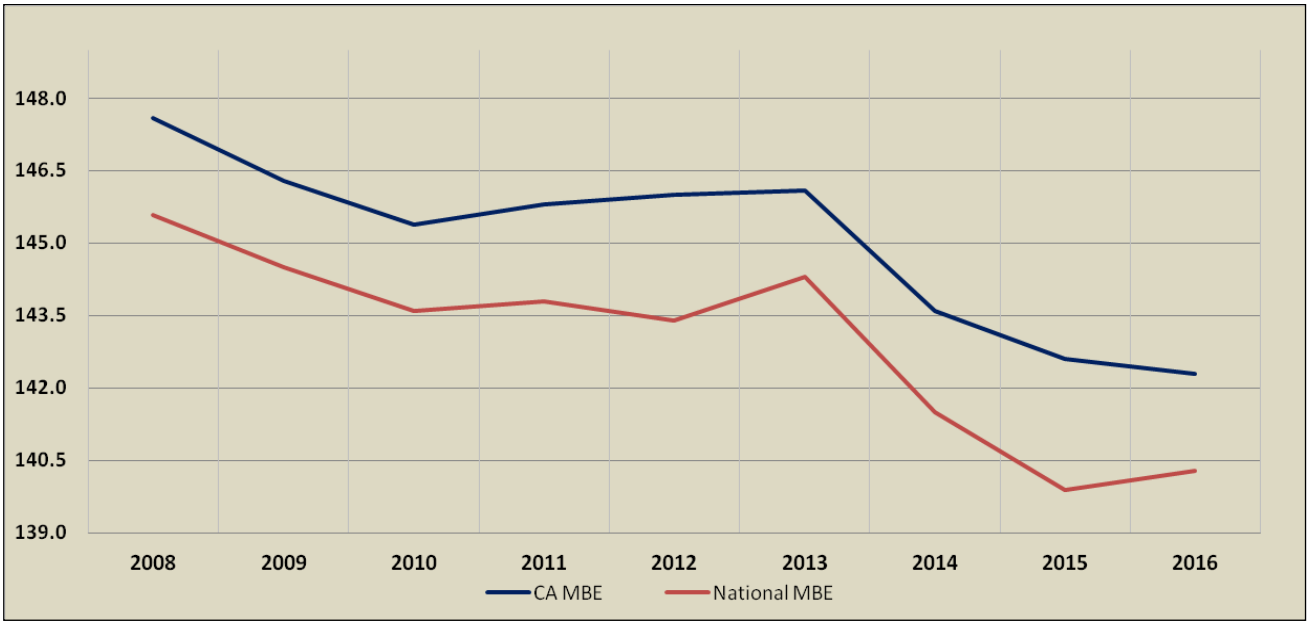
Finally, this study did not address whether the content of the CBE remains relevant to assessing the minimum competency to practice law, or whether the current standard remains appropriate in today's practice environment. These are issues that would also require different data and study methods.

I. BACKGROUND

In recent years, there has been a fairly steady decrease in the passing rate on the California Bar Examination (CBE). From its recent peak in July of 2008 to the most recent 2016 July administration, the percentage of applicants passing the exam has fallen by 18% (from 62% to 44%). The decline has been a steady one. During no 8-year period since the examination was in its current configuration has the passing rate decreased by this amount. This downward trend mirrors a similar pattern observed in the average Multistate Bar Examination (MBE) performance in California over the same period; a drop from 1476 to 1423. It is worthwhile noting that the trends observed in California are consistent with those observed nationally over the same period (see Figure 1).

Figure 1

Comparison of California and National Average MBE Performance 2008 through 2016 July Administrations



Much has been written recently about possible causes for these drops. Some have theorized that the test takers themselves have changed. The National Conference of Bar Examiners (NCBE), authors of the MBE, has published several pieces suggesting this and attesting to the continuing psychometric strengths of the exam. They point to the changing landscape of legal education reflected in lowered admission numbers, a decline in the quality of the applicant pool, and shifting attrition and transfer policies. This argument has been somewhat corroborated in statistics reported by the American Bar Association. For entering law school classes of 2005 and 2013[1], the number of law school applications fell 38% (from 95,800 to 59,400), the number of admissions dropped by 19% (from 56,100 to 45,700) and eventual matriculations decreased by 17% (from 45,800 to

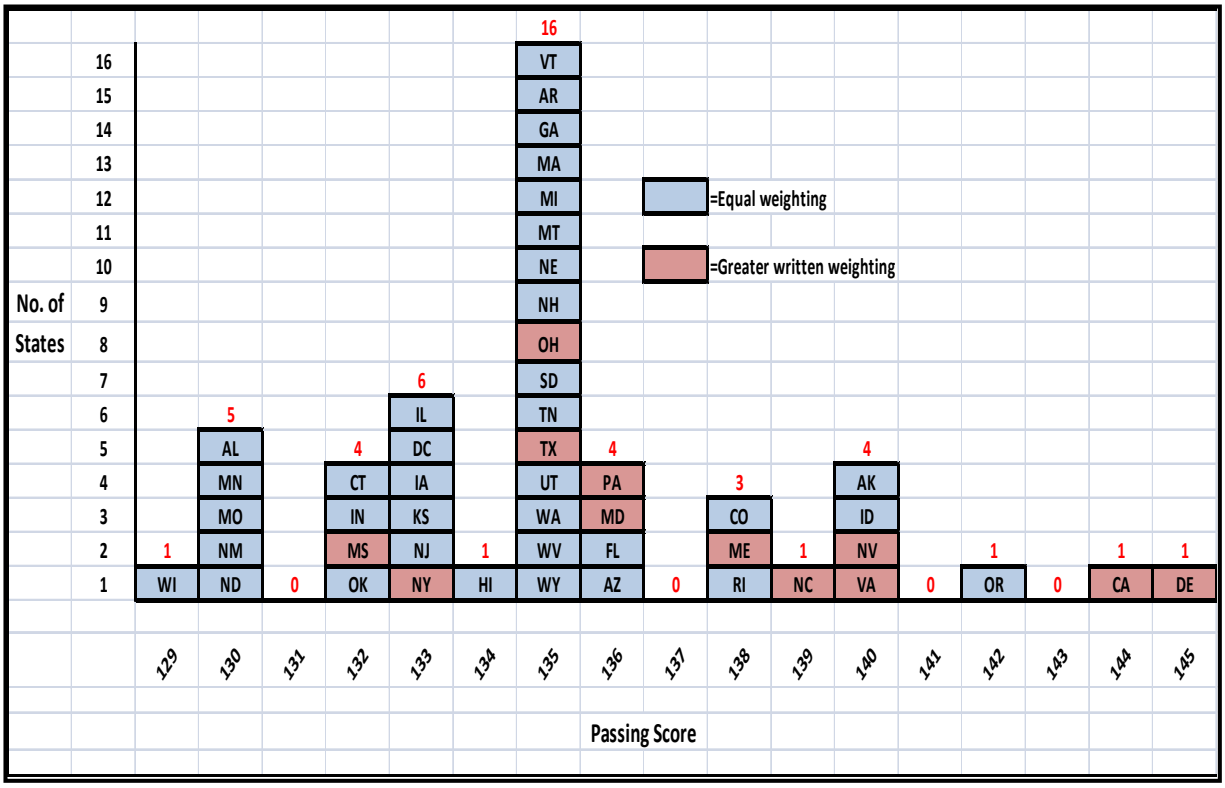
[1] These classes would have made up the majority of first time test takers sitting for the July 2008 and 2016 bar examinations.

37,940) with no corresponding decrease in the number of schools during that period. As a result, a much higher percentage of applicants were admitted to ABA schools in 2013 (75%) than in 2005 (59%).

Law school deans have rebutted these arguments. They have posited that the examination itself has gotten more difficult. The deans have questioned whether it remains an appropriate measure of minimal competency to practice law and whether the content is relevant. They also argue that the standard that is being applied in determining minimal competency to practice in California has been set too high, and should be more in line with other states (see Figure 2). They point to the increased curricular emphasis and instructional time that has been placed on bar preparation skills and legal analysis in recent years. Finally, the deans suggest that the average credentials (LSAT scores and Undergraduate GPA) have generally not declined, and where they have, they are in no way decreasing at the rates that their bar passage has.

Figure 2

The Passing Scores
States Using the Multistate Bar Examination
(in MBE Units)



II. PURPOSE OF THIS STUDY

A shift over time in performance on an examination such as the CBE is generally a function of one or more of three possible reasons: (a) the examination itself has changed in some manner, (b) the overall ability or preparedness level of the applicants sitting for the exam has changed, and/or (c) the composition of the test-taking population has been altered. In response to the ongoing public debate and to help untangle the relative impact of these causes, the California Committee of Bar Examiners (The Committee) requested that an initial study be conducted.

This initial study would be limited to analyses of existing, readily available electronic Admissions/Examination Results databases. The purpose of the study would be to (1) establish a statistical baseline to profile the changes that have occurred; and (2) determine whether insights could be provided into factors that might have contributed to the decline. The study would both draw on existing broad statistical summaries and technical reports prepared after each examination, and supplement them with additional, more detailed analyses of the electronic databases that would focus on year-over-year changes.

Thus in summary, the primary objectives of this study were to organize and investigate historical databases for the purpose of establishing a baseline for the changes that have occurred over time, and to investigate any emerging patterns that that could shed light on any or all of the three potential reasons for the decrease in scores and passage rates.

III. METHODS

A. Study Data

The Bar Admissions Office of the State Bar ("Admissions") maintains a base of information for each applicant who sits for the CBE. In additional to basic demographic information (e.g., gender and race/ethnicity), the applicants' scores on each section of the examination and final pass/fail disposition are maintained for all applicants. For the current study, we focused on three specific administration years:

- 2008, a period when bar passage rates were at their highest in recent history
- 2016, the most recent period when scores and bar passage rates have been at their lowest since at least 1990
- 2012, a midpoint between the two years when scores were on the decline

We reasoned that if patterns did exist, they would come to light by focusing on the most recent years with the most extreme differences.

Additional factors contributed to the selection of these periods. The CBEs during these years shared the following similarities[2] in that:

[2] The configuration and scoring of the CA Bar Examination has changed over the years. We reasoned that it would best to eliminate exams from those periods so as to insure apple-to-apple comparisons.

- They were configured the same(i.e., the MBE, 6 Written Essays, and 2 Performance Tasks (PT))
- They were scored the same (i.e., Each PT was given 2 times the weight of an Essay)
- They were scaled the same (i.e., the Raw Written Score was scaled to the mean and sd of the MBE)
- Total Scale Scores (TSS) were calculated in the same manner (i.e., .35 x MBE + .65 x Written)
- Phase II regrading score bands were the same (i.e., 1390-1439.99)[3]
- The score required for passing remained the same (i.e., 1440[4])

For each applicant testing within those years, we extracted the following demographic and performance data[5]:

1. Racial/ethnic status
2. Gender
3. Applicant's law school designation (which includes other non-traditional designations such as foreign trained)
4. Number of examination attempts at the time of administration
5. Attorney applicant status
6. Examination administration (February vs. July)
7. MBE Scale Score
8. Written Scale Score
9. Total Scale Score (TSS)
10. Pass/Fail disposition

Standard reporting of each CBE's general statistics routinely re-categorizes the 300+ law school designations into more homogenous clusters. Previous analyses have found that average examination performance between these clusters varies significantly. Therefore, to facilitate analysis and reporting, we established similar clusters. They included:

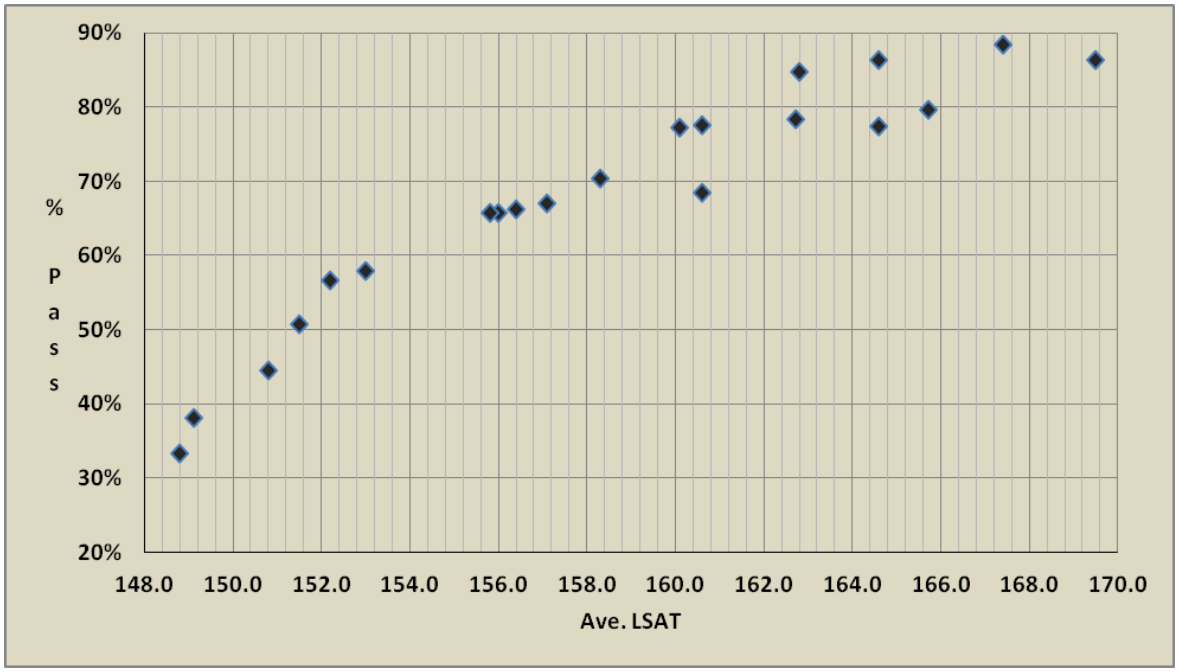
- California ABA-Approved Institutions (CA-ABA)
- Non-California ABA-Approved Institutions (NCA-ABA)
- California Accredited Institutions (ACC)
- California Unaccredited Institutions (NAC)
- Foreign Trained (FOR)

Further Classification of CA-ABA Schools. Past research has identified wide diversity in examination performance between students from the various CA-ABA institutions and found that these differences were highly correlated with the Average Law School Admission Test (LSAT) scores at these institutions (see Figure 3). As a result, we reasoned that it would be valuable to further categorize these schools into more homogeneous groups in a search for deviations in patterns of performance.

[3] Between 2008 and 2012, the Phase III scoring process changed somewhat, but should have no impact on the analyses proposed here given the relatively small number of applicants experiencing this scoring.
[4] CA multiples the MBE by a factor of 10. Thus, the 1440 is equivalent to 144 on the original MBE scale.
[5] Only applicants completing all sections of the CBE were chosen.

Figure 3

**Relationship between Average Law School LSAT and Bar Passage Rates*
1998-2007**



* **Each point on the graph represents the 10 year average LSAT and Passing Rate for students from one school**

Since LSAT scores are no longer collected during the application and admissions process, we looked to an outside source[6] for these data. The website lawschooltransparency.com provided median LSAT scores for each CA-ABA school. Scores were available for 2010 and 2014. We used the 2010 results[7] since those statistics would most closely reflect the class which took the midpoint examination in the study. We attempted to establish roughly equal number of law schools in each group and find a break point in the Median LSAT for the grouping. Our analysis resulted in the following groups:

- Level I - 7 schools; Median LSAT Range (150-155)
- Level II - 6 schools; Median LSAT Range (158-161)
- Level III - 8 schools; Median LSAT Range (163-170)

[6] The last year that individual applicant LSAT scores were collected during the admissions process was 2007
[7] The correlation between 2010 and 2014 LSATs was .96 suggesting little change in the relative standings of the schools over time

B. Research Questions

Given the available data, this study sought to address five (5) major research questions.

1. ***How has the composition of the test-taking population changed over time?***
   a) How has the absolute number of test takers differed?
   b) What changes have occurred in the relative "mix" of test-takers, i.e., do certain historically lower performing groups now make up a higher proportion of the test-taking population?

As the number of applications to law schools have decreased, it is possible that characteristics of students (measureable or otherwise) have changed over time. Historically, selected applicants from certain subgroups have performed more poorly on the bar examination and passed at a lower rate than others (e.g., NAC vs CA-ABA schools). If the test taking population as a whole is more "saturated" with these lower performing groups, it might be one cause for decreasing scores.

2. ***To what degree have examination scores & final pass/fail disposition changed over time?***
   a) Has the magnitude of the changes been consistent across sections of the examination?
   b) Have each of the relevant sub-groups experienced similar changes, or have some groups experienced greater changes than others?

The simple change in the overall passage rate is a gross statistic. Knowing if specific groups of applicants experienced larger or smaller decreases in performance is essential to a gaining an understanding of the change. Additionally, pass/fail disposition is based upon actual examination scores; a closer examination of the size of differences is essential.

3. ***To what degree has the shape of the distribution of scores changed, i.e. while the mean scores have changed, have other attributes (e.g., the median, relevant quartiles, etc) shifted as well?***

At this point, it is unclear whether the change in scores leading to the decreasing passage rate is consistent throughout the score distributions, or more heavily concentrated in one or more locations (e.g., close to the passing standard). It is reasonable to determine what size of improvement in performance on recent examinations would have led to increased passage rates.

4. ***Has the likelihood of eventually passing (e.g., after 2 years) changed over time?***
   a) Are individual examinees who must repeat the exam more or less likely to pass upon retaking the exam?
   b) Has perseverance of failing examinees remained consistent?
   c) What do these patterns look like by relevant sub-groups, and how have they changed?

Preliminary evidence suggests that the recent passing rates have not dropped as rapidly for repeating applicants (those taking the exam for a 2nd, 3rd, or more time) as in the past. This may imply that an "eventual" bar passage rate may not have changed as drastically as the first time rate. What we may be seeing is that recent applicants are taking longer to pass. If this were the case, it may point to possible changes in applicants' bar preparation that have occurred since 2008.

Analyses of the success of test repeaters require that a given applicant be tracked over time, i.e., longitudinally. To compile longitudinal data for the purposes of this study, we first obtained data for the cohort of students who took the exam for the first time in July 2008, and followed them forward through February 2010, a total of four examinations. For a second cohort who had taken the exam more recently, we selected students who first sat for the CBE in July 2014 (when the passing rate first dipped below 50%) and followed them forward through February 2016, an additional four examinations.

5. Have other statistical/psychometric properties of the examination changed over time in such a way to impact applicant scores?

a) Has the reliability of the overall examination or its individual sections changed?
b) Has the nature of the relationship of the sections changed? For example, if historically applicants performed similarly on specific sections of the exam, either doing well or poorly on both sections, has that pattern persisted?

The amount of measurement error that exists in applicant scores is a function of the reliability of the respective sections (i.e., written and MBE) and the degree of relationship between them. Reliability is a measure of the degree of stability or consistency of scores on a test and is one of key indicators of a test's psychometric properties.[8] The lower the reliability, the higher the amount of error that exists in the measurement. Overall reliability on the CBE itself is a function of the separate reliabilities of the Written section, the MBE and the degree of correlation between the two. As any of these three values change, so does the reliability.

6. How would bar passage rates change if the cut point were set at a standard used by other states?

a) What would the passage rates have been if a different passing score had been established?
b) Would the decline in passage rates during the study timeframe been as pronounced under such a circumstance?
c) Would any relevant sub-group have seen larger increases or decreases than others?

The California standard (i.e., 1440) for passing the CBE did not changed over the 9 year time frame of our analyses. As previously discussed, this standard is the second highest in the country, and questions have been raised as to whether bar passage rates would have declined as steeply if the standard was lower. An additional related question was

[8] Validity is another major psychometric property of a test. Data available to this study precludes an evaluation of any changes that may have occurred since 2008 in any of the various measures of validity that are used.

how the decreases in passing rates under an alternative standard would compare to that of a similarly situated state (i.e., one of comparable size, applicant composition and passing standard).

IV. RESULTS

This section discusses the data analyses and outcomes relevant to each of the study research questions identified above. For the most part, we present findings for both July and February administrations. For some analyses we present results for July only since applicants sitting for this administration generally are more representative of the typical recent law school graduate. All calculated statistics are presented in the tables but only key findings (e.g., significant differences between CBE years or subgroups) are discussed in the text.

1. How has the composition of the test-taking population changed over time?

Figures 4 and 5 illustrate the trend in applicants sitting for the July and February CBE since 1990.

Figure 4

Number of Examinees Taking July CBEs
1990 through 2016

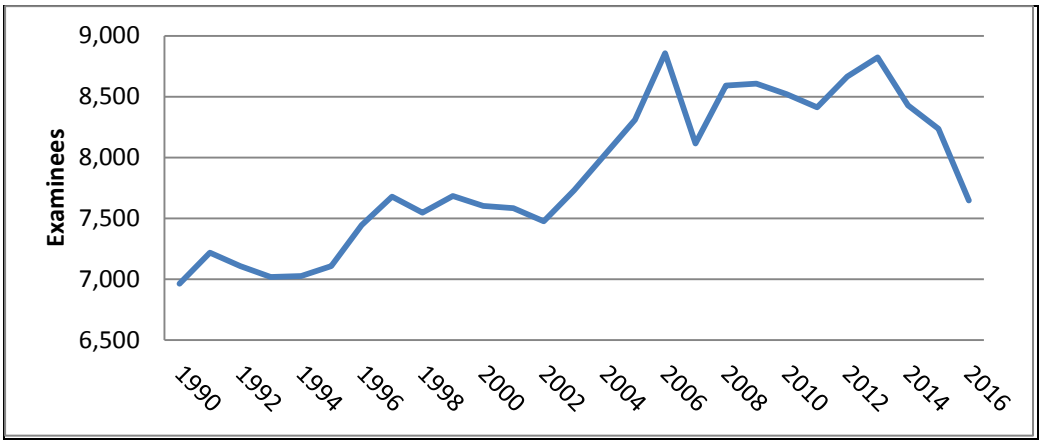


Figure 4 depicts the gradual rise in July examinees peaking in 2006 and again in 2013, and beginning a sharp drop in 2014. The February counts (Figure 5) have tended to track with those of the July examinations, though the downward trend seen in the July counts during the past two has been countered by an upward trend in the number of February test takers. This uptick may be a function of more applicants repeating the examination.

Figure 5

Number of Examinees Taking February CBEs
1990 through 2016

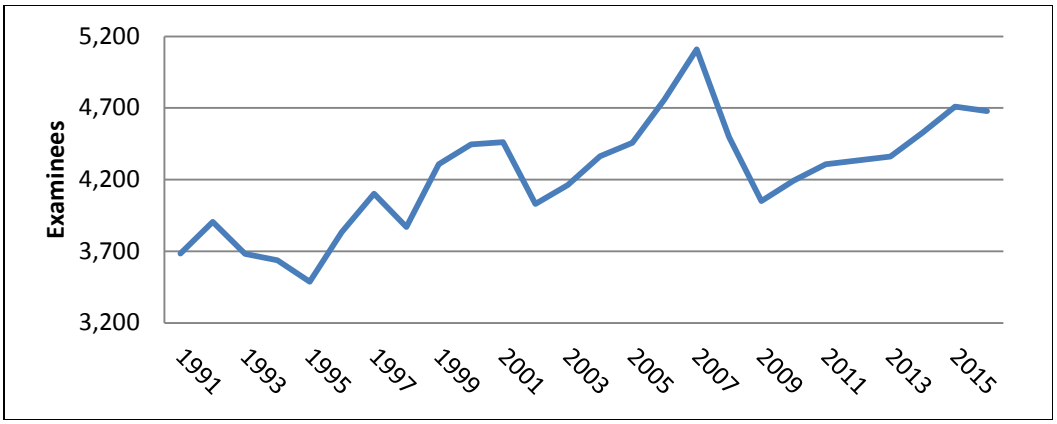


Table 1 provides the characteristics of applicants sitting for examinations in each of the three years included in our study time frame (2008, 2012 and 2016). Since the populations of test-takers for the July and February bar administrations have traditionally varied in terms of size and composition, we examined differences within each administration separately. These point-in-time snapshots show small, but interesting variances in the composition of the applicant populations in each year.

For the July CBE, we note that:

- In comparison to 2008, there were 11% fewer applicants in 2016. This is the largest change during any 8-year testing period since 1990.

- In comparison to 2008, the 2016 population of test-takers included a slightly higher proportion of minority applicants, notably Hispanics (5%) and a corresponding lower proportion of White applicants (6%)[9]. As discussed later, minority applicants have tended to have lower scores and passage rates than Whites.

- There were 5% fewer first time takers in 2016 than in 2008 (72% compared to 67%), and a corresponding 5% increase in the proportion of test repeaters. First time applicants traditionally have performed higher than those repeating the CBE.

[9] A small number of applicants do not report their race/ethnicity or report as some other group. The percentages are based only on applicants in the four major groups.

Table 1

Composition of the CBE Applicant Pool in 2008, 2012, 2016
July and February Administrations

Metric	July CBE				February CBE			
	Year			Change	Year			Change
	2008	2012	2016	2008-2016	2008	2012	2016	2008-2016
Examinees	8,590	8,664	7,648	-11%	4,497	4,334	4,678	4%
School								
CA ABA	53%	55%	53%	0%	40%	39%	38%	-2%
Level I	29%	28%	31%	2%	50%	47%	48%	-2%
Level II	36%	35%	31%	-5%	30%	34%	30%	0%
Level III	34%	35%	37%	3%	18%	18%	21%	3%
Non CA ABA	22%	20%	18%	-4%	17%	18%	17%	0%
CA Accredited	9%	7%	10%	1%	13%	12%	14%	1%
CA Non-Accredited	3%	2%	2%	-1%	7%	6%	4%	-3%
Foreign	3%	3%	5%	2%	5%	6%	7%	2%
Exams Taken								
1st	72%	74%	67%	-5%	33%	33%	29%	-4%
2nd	7%	7%	9%	2%	33%	33%	38%	5%
3rd	7%	6%	10%	3%	8%	9%	10%	2%
> 3rd	12%	11%	13%	1%	24%	23%	22%	-2%
Racial/Ethnic								
Asian	18%	18%	20%	2%	18%	19%	21%	3%
Hispanic	9%	10%	14%	5%	11%	12%	14%	3%
Black	5%	5%	6%	1%	8%	7%	7%	-1%
White	57%	56%	51%	-6%	52%	51%	49%	-3%
Gender								
Male	52%	53%	48%	-4%	54%	51%	50%	-4%
Attorney	9%	9%	11%	2%	17%	17%	19%	2%

* Multi-group categories may not add to 100% due to missing information or small numbers in an "other" group

- 53% of applicants graduated from CA-ABA schools in 2008 and this did not change in 2016. However, there were 4% fewer students coming from NCA-ABA schools (22% versus 18%). Foreign trained applicants, a traditionally low performing group, increased slightly from 3% to 5%.

- By 2016 males no longer made up the majority of examinees (48% vs 52% females). Given the historical similarities in scores between the gender groups, this change would be estimated to have minimal impact.

With respect to the February administrations, we observed:

- The number of applicants for the February 2016 administration was 4% greater than the number in 2008 (4,678 versus 4.497) but the relative percentage of first time takers was reduced by 4% (from 33% to 29%). This could suggest that the recent decrease in the July passage rates may be "feeding" additional applicants into the February administrations.[10]

- Similar to the July examinations, the proportion of Asian and Hispanic examinees was higher in 2016 than 9 years earlier while the percentage of Whites was 3% lower.

- The proportion of students from Level III (high LSAT) schools was 3% greater in 2016 than in 2008.

- The proportion of attorney applicants in the February exam was higher than in the July exam in both 2008 and 2016, and for both administrations in both years the proportion of attorneys sitting for the bar was 2% greater.

Table 1 shows that the proportion of applicant groups that have historically scored lower on the CBE was somewhat greater in 2016 than in 2008. A full evaluation of any relationship between these changes in the composition of the applicant population and a reduction in scores requires addressing the remaining research questions.

2. To what degree have <u>examination scores</u> & final <u>pass/fail disposition</u> changed over time?

<u>Total Population.</u> Table 2 presents information on the average performance on each section of the exam and the total scores (expressed in scale score points), along with the percentage passing the examination in the three years under study. Data is presented for both the July and February CBE.

Inspection of the table reveals that while the absolute change in the MBE and Written sections of the examination have differed, the percentage decreases in scores are equal (4% for July and 1% in February). *This result indicates that across all applicants, no one section of the examination is contributing to the decrease in passing rates more than another.*

[10] The "tracking/persistency" portion of the analyses presented later will shed more light on this issue

Table 2

Average CBE Performance & Bar Passage Rates By Administration

Year	July					February				
	N	Ave. MBE	Ave. Written	Ave. Total	% Pass	N	Ave. MBE	Ave. Written	Ave. Total	% Pass
2008	8,590	1476	1481	1479	62%	4,497	1405	1400	1402	40%
2012	8,664	1460	1456	1457	56%	4,334	1407	1407	1407	43%
2016	7,648	1423	1415	1418	44%	4,678	1388	1387	1387	36%
2008-2016										
Diff.	-942	-53	-66	-61	-18%	181	-17	-13	-15	-4%
% Change	-11%	-4%	-4%	-4%	-29%	4%	-1%	-1%	-1%	-10.0%

It further suggests that whatever the different skills being measured on the respective parts of the test, all have decreased at a similar pace. Overall, the average Total Scale Score (TSS) has dropped 61 scale score points in July (from 1479 to 1418) and 13 points in February (from 1400 to 1337). By way of reference, in 2016, the average score actually fell below the passing standard of 1440.[11] In terms of standard deviation (Sd) units, this represents slightly less than a ½ Sd change in July and a 10% Sd change in February.

The TSS drop was accompanied by a corresponding decrease in passing rates for the July exams; there was a steady decline in these rates from 62% to 56% to 44% in 2008, 2012 and 2016, respectively. The change in passing rates in February, however, rose between 2008 and 2012 (from 40% to 43%), followed by drop to 36% in 2016.

We next examine whether different segments of the applicant pool experienced differing degrees of change from 2008 to 2016. Given the substantial difference between July and February administrations, we present findings for the July examinations. Where findings are significantly different for February administration, we point these out.

Repeater Status. Table 3 presents similar data to Table 2, stratified by whether applicants were sitting for the first time ("first timers"), or repeating the examination ("repeaters"). As known from historical results, first timers perform consistently higher than repeaters and that fact is illustrated in Table 2. The gap in TSS between the two groups in 2008 was 150 scale score points (a full Sd.); however, that gap decreased on average to 137 points by 2016.

[11] This situation recently began in 2013 when the mean score fell to 1436 and has occurred in three other administrations since 1990.

Table 3

Average CBE Performance & Bar Passage Rates By Repeater Status
July Administration

`	1st Time Taker				Repeater			
	Ave. MBE	Ave. Written	Ave. Total	% Pass	Ave. MBE	Ave. Written	Ave. Total	% Pass
2008	1515	1523	1520	75%	1373	1368	1370	28%
2012	1493	1495	1494	69%	1365	1340	1349	18%
2016	1458	1461	1460	57%	1353	1323	1333	17%
2008-2016								
Diff.	-57	-62	-60	-18%	-20	-45	-37	-11%
% Change	-4%	-4%	-4%	-24%	-1%	-3%	-3%	-39%

First timers experienced similar rates of decrease in their MBE and Written scores, while repeaters' MBE scores dropped by 1% as compared to a 3% drop in their Written scores. The absolute decrease in passing rates for first timers between 2008 and 2016 (18%) followed the pattern for the entire test taking pool, while the absolute decrease for repeaters was almost half that amount (11%). Additionally, for repeaters the sharpest decrease was seen in 2012 (18%; a drop of 10% from 2008). The change from 2012 to 2016 was only 1% as compared to the 8% decrease for first timers. This pattern tends to suggest that the recent applicants repeating the examination may have been qualitatively different than their predecessors.

Law School. Table 4 presents change in scores and passing rates stratified by the type of school that the applicant attended. Historically, average scores and passing rates have been highest among the CA-ABA and NCA-ABA schools. Scores and passing rates for ACC and NAC have been much lower. For example, in 2008 the passing rate at CA-ABA schools was 74% as compared to 21% at NAC schools (a net difference of over 50%). Yet, in terms of score changes and decreases in passage rates over the study time frame, students from CA-ABA schools had the largest absolute changes in scores and bar passage rates.

Table 4

Average CBE Performance & Bar Passage Rates by Type of Law School Attended
July Administration

Year	CA ABA				Non CA ABA				CA Accredited				CA Unaccredited			
	Ave. MBE	Ave. Written	Ave. Total	% Pass	Ave. MBE	Ave. Written	Ave. Total	% Pass	Ave. MBE	Ave. Written	Ave. Total	% Pass	Ave. MBE	Ave. Written	Ave. Total	% Pass
2008	1510	1525	1519	74%	1497	1494	1495	66%	1355	1367	1363	26%	1375	1341	1353	21%
2012	1491	1503	1499	69%	1476	1452	1461	55%	1350	1347	1348	19%	1369	1313	1332	13%
2016	1457	1460	1459	54%	1448	1434	1439	48%	1329	1312	1318	13%	1350	1280	1304	14%
2008-2016																
Diff.	-53	-65	-60	-20%	-49	-60	-56	-18%	-26	-55	-45	-13%	-25	-61	-49	-7%
% Change	-4%	-4%	-4%	-27%	-3%	-4%	-4%	-27%	-2%	-4%	-3%	-50%	-2%	-5%	-4%	-33%

Average scores by section dropped equally (roughly 4%) in both in and out-of-state ABA schools, while students from the ACC and NAC experienced greater drops in their Written sections (4% and 5% respectively) than on the MBE (2%). Correspondingly, the absolute drop in bar passage rates was greater for students from the ABA schools (20% and 18%) than in the non-ABA schools (13% for ACC and 7% for NAC). As shown at the bottom of Table 4, the *absolute drop* in the passing rate is quite different from the *percentage change* in the passing rate. For example, while the passing rates for students in ACC schools dropped by only 13%, that drop represented a 50% decrease from the 26% level in 2008.

When we look more deeply into the changes in performance of students from CA-ABA schools, some interesting trends begin to emerge. Table 5 provides data on the performance of applicants from schools based upon the median LSAT for students at those schools. Both average section scores and passing rates for the Level III schools (i.e., those with the highest median LSAT scores) were the highest of the three school groups in 2008 and remained that way in 2016. Students from Level III schools also showed the smallest decrease in passage rates at 11%, and the smallest and most consistent changes in examination section scores (3%) and overall TSS. For applicants from schools with lower median LSAT scores (Levels II and I), the decrease in performance between 2008 and 2016 is much greater.

Table 5

Average CBE Performance & Bar Passage Rates by CA ABA Law School Level
July Administration

Year	Level I LSAT Schools				Level II LSAT School				Level III LSAT Schools			
	Ave. MBE	Ave. Written	Ave. Total	% Pass	Ave. MBE	Ave. Written	Ave. Total	% Pass	Ave. MBE	Ave. Written	Ave. Total	% Pass
2008	1455	1476	1468	61%	1513	1529	1523	77%	1553	1562	1559	83%
2012	1439	1460	1453	55%	1480	1497	1491	69%	1542	1543	1543	81%
2016	1389	1390	1389	32%	1462	1458	1460	56%	1509	1519	1516	72%
2008-2016												
Difference	-66	-86	-79	-29%	-51	-71	-63	-21%	-44	-43	-43	-11%
% Change	-5%	-6%	-5%	-48%	-3%	-5%	-4%	-27%	-3%	-3%	-3%	-13%

While not all students in the various law school groupings share an identical LSAT score, their LSAT does tend to be more similar to those in their own school group than students attending schools from other Levels. Although there have been decreases in CBE performance in all Levels in recent years, the fact that the changes are more pronounced in the Levels I and II schools may suggest that the quality (and possible ability level) of students from those schools have changed at a more rapid pace than students from the Level III schools.

Racial/Ethnic Group. Table 6 presents similar statistics for the July examinations stratified by racial/ethnic group. Historically, White students have made up the majority of students sitting for the CBE and have had the highest scores and bar passage rates. When we look at how CBE performance has changed by racial/ethnic group over the study time frame, we see that Whites have tended to behave similarly to the various minority groups. Mean Written scores have dropped by 4% between 2008 and 2016, which is exactly the pattern seen in Blacks and Hispanics. Scores for Asians, a group whose ranks have proportionately increased since 2008, dropped by 1% more. Across all ethnic groups, TSS have decreased by either 3% or 4%, and the decrease in bar passage rates differ by only have 5% between the groups (18% in Asians, 17% for Whites, 15% for Hispanics and 13% for Blacks). The largest relative decrease in passage rates was experienced by Blacks where their change from 34% in 2008 to 21% in 2016 represents an overall 38% decrease (compared to 32%, 31% and 24% for Asians, Hispanics and Whites, respectively).

Table 6

Average CBE Performance & Bar Passage Rates by Racial/Ethnic Group
July Administration

Year	Asian				Hispanic				Black				White			
	Ave. MBE	Ave. Written	Ave. Total	% Pass	Ave. MBE	Ave. Written	Ave. Total	% Pass	Ave. MBE	Ave. Written	Ave. Total	% Pass	Ave. MBE	Ave. Written	Ave. Total	% Pass
2008	1447	1462	1457	56%	1427	1437	1433	49%	1389	1382	1384	34%	1504	1505	1504	69%
2012	1431	1443	1439	51%	1411	1414	1413	42%	1380	1362	1370	28%	1491	1480	1484	64%
2016	1383	1396	1392	38%	1388	1381	1383	34%	1361	1327	1339	21%	1459	1445	1450	52%
2008-2016																
Difference	-64	-66	-65	-18%	-39	-56	-50	-15%	-28	-55	-45	-13%	-45	-60	-54	-17%
% Change	-4%	-5%	-4%	-32%	-3%	-4%	-3%	-31%	-2%	-4%	-3%	-38%	-3%	-4%	-4%	-25%

Gender. Finally, Table 7 presents performance statistics by gender. Both CBE scale scores and bar passage rates have historically been fairly equal. In 2008 the there was only a 9 point difference between males and females and 2% difference in passage rates (females higher in both cases)[12]. In 2012, performance was identical for males and females, while in 2016, the female passing rate was 1% lower for females than male examinees (43% vs 44%). This slight shift is evidenced in the 2008 to 2016 *% Change* data showing a net decrease in pass rates of 20% for females and 17% in males.

[12] A pattern has existed for many years whereby female test takers score more highly on the Written section of the CBE while the reverse is true for the MBE. Interestingly the gap has widened on the MBE while narrowing slightly on the Written section.

Table 7

Average CBE Performance & Bar Passage Rates by Gender
July Administration

Year	Females				Males			
	Ave. MBE	Ave. Written	Ave. Total	% Pass	Ave. MBE	Ave. Written	Ave. Total	% Pass
2008	1462	1496	1484	63%	1489	1467	1475	61%
2012	1439	1467	1457	56%	1479	1446	1458	56%
2016	1403	1421	1415	43%	1443	1409	1421	44%
2008-2016								
Difference	-59	-75	-69	-20%	-46	-58	-54	-17%
% Change	-4%	-5%	-5%	-32%	-3%	-4%	-4%	-28%

"Multi-Characteristic" Estimation Model. The preceding tables have shown some changes in the composition of the CBE applicant population over the study period (Table 1), along with changes in CBE performance by individual characteristics of applicants including repeater status, type of law school, race/ethnicity and gender (Tables 3-7). An applicant however is some combination of these individual attributes. For example, they may be a Hispanic female coming from a Level III ABA school who repeated the exam for the second time, or a White male who graduated from an accredited law school making their first attempt. Additionally, the combination of characteristics represented by applicants in each year's test-taking population varies over time.

To estimate the impact that that the change in applicant mixes from 2008 to 2012 and 2016 may have had on performance in the latter two years, we developed an estimation model. In the model we calculated the bar passage rates and average TSS in 2008 for all combinations of number of exams taken (first time vs. repeater), law school type (including the separate CA-ABA Levels), racial/ethnic group, and gender. We then applied those statistics to the applicants in the same groups in 2012 and 2016, re-weighted them based upon the applicant counts in the respective groups, and recalculated (i.e., estimated) the overall mean TSS and bar passage rates. The results are summarized in Table 8.

Results from Table 8 shows that the changed composition of examinees would have led to reduced performance in both 2012 and 2016, all other things held equal. For the July administration in 2016, the TSS would have been expected to drop by 12 points (1479 – 1467) and the passing rate expected to drop by 3% (62% - 59%). The actual decreases for both measures were much greater, however: a 60 point decrease in the TSS and an 18% decline in the passing rate. *The results suggest that for the July administration only 20% of the change in TSS (12/60) and 17% of the change in passage rates were due to the shift in applicant mix.*

Table 8

Projected vs. Actual CBE Performance

Year	Average Total Score		% Passing	
	Projected	Actual	Projected	Actual
July				
2008		1479		62%
2012	1483	1457	64%	56%
2016	1467	1419	59%	44%
2008-2016 Difference	-12	-60	-3%	-18%
February				
2008		1402		40%
2012	1402	1407	40%	43%
2016	1399	1387	40%	36%
2008-2016 Difference	-3	-15	0%	-4%

In February, performance would have been estimated to have dropped slightly as well, but not to the same degree as July. For example, the 2016 pass rate would have been estimated to remain exactly the same as in 2008 (as compared to an actual drop of 4%), while Average TSS would have been estimated to have changed by only 3 points (20% of the actual change). *These findings strongly suggest that there are other, unmeasured characteristics in the population of test takers and/or the testing that has led to the observed declines in passage rates between 2008 and 2016.*

3. How has the distribution of scores changed, i.e. while the mean scores have changed, have other attributes (e.g., the median, relevant quartiles, etc) changed as well?

Often the focus on a simple measure of central tendency (e.g., a mean) masks other interesting information in large samples such as that for the thousands of applicants sitting for the CBE. While the previous tables showed that the average scores have trended downwards from 2008 to 2016, they don't indicate where the changes have occurred in the distribution, nor how. For example, average scores by themselves will not indicate whether large amounts of applicants have scored just below the passing standard of 1440, while a second large cluster of test-takers with much lower scores led to an observed decline in the "average" test score.

We explore differences in the score distributions for 2008 and 2016 below. Since the previous data has suggested that more significant changes have occurred in the July administrations, results in this section are reported for those examinations only.

Distribution Similarities and Differences. Table 9 reports the TSS scores associated with various percentiles within the distributions of the 2008 and 2016 examinations. A percentile is defined as the percentage of observations (i.e., applicants) scoring at or below the given score. Table 9 presents data on 5 key percentile points: the three "quartiles" which are the 25th percentile, the 50th percentile (i.e. the median or midpoint), and the 75th percentile; the 10th percentile which is located at the bottom or the distribution and the 90th percentile, which is located at the top of the distribution. In addition to the TSS, we report this data for both the Written and MBE sections.

As can be seen in Table 9 the scores associated with each percentile point for each scale score are lower in 2016 than 2008, though the sizes of the differences are not consistent across the percentile points or by examination section. For example, with respect to the MBE, we see that the bottom 10% of the 2008 applicant pool scored a 1267 or higher as compared to the bottom 10% in 2016 scoring only 1197; a 70 point difference (almost ½ Sd). It can also be seen that as one moves up the distribution, the sizes of the difference begin to get smaller (the 90th percentile in 2008 was 1673 compared to 1631 in 2016, a difference of only 42 points). This finding suggests that, in comparison to 2008, a greater proportion of the lower performing students (on the MBE) in 2016 clustered at the bottom of the distribution for that year. The pattern is slightly changed on the Written section where differences appear to be more consistent in the middle ¾'s of the distribution and slightly smaller at the tails.

Table 9 also reports the size of the standard deviation or the measure of score spread. On average, performance scores have a greater spread in 2016 on both sections and overall. The cause of this additional score spread cannot be determined from the available data, but it does suggest potentially greater variation in the applicant pool.

Table 9

Total Scale Scores at Various Percentile Points
on the 2008 and 2016 CBEs
July Administrations

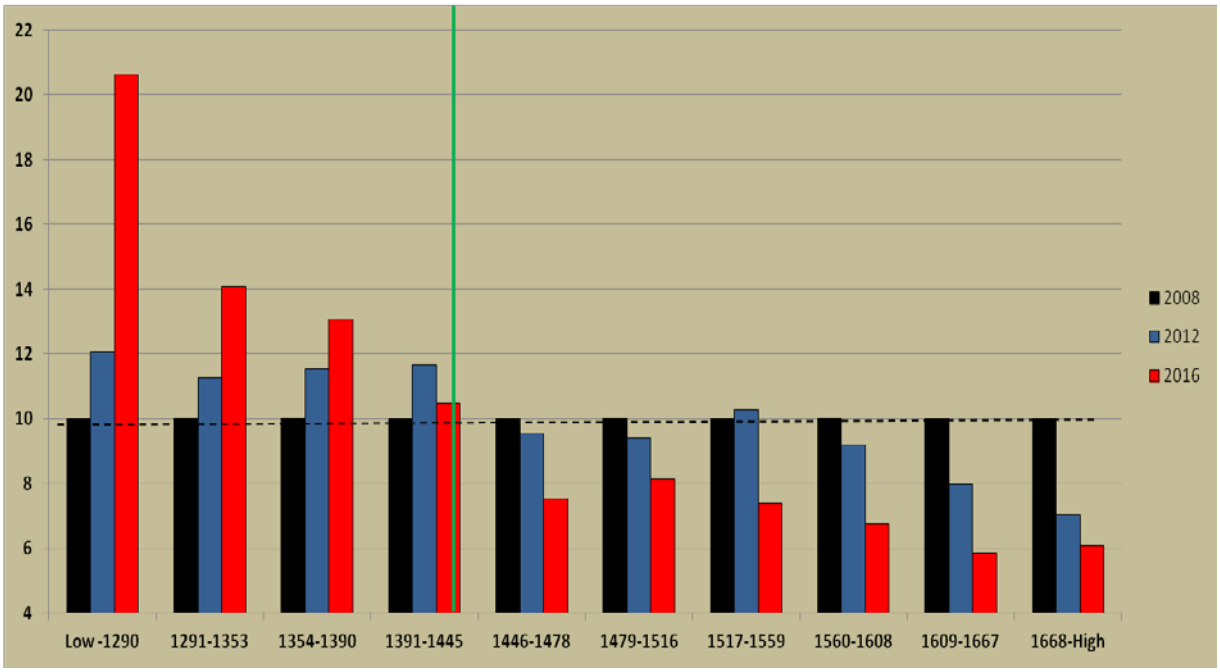
Distribution Points	MBE Score			Written Score			Total Score		
	2008	2016	Diff	2008	2016	Diff	2008	2016	Diff
10th Pctl	1267	1197	-70	1282	1220	-62	1292	1227	-65
1st Quartile	1375	1315	-60	1364	1290	-74	1374	1313	-61
Median	1487	1437	-50	1473	1394	-79	1478	1402	-76
3rd Quartile	1593	1543	-50	1595	1516	-79	1582	1522	-60
90th Pctl	1673	1631	-42	1689	1638	-51	1667	1627	-40
Std Dev	155	167	12	158	165	7	145	155	10

"Exploring the Tail". The increase in score spread and the size of the difference at the 10th percentile of the MBE (an equated measure and the more reliable of the two sections) between 2008 and 2016, led to further exploration of possible explanations for the observed differences between the two periods.

To make a direct comparison we first established the deciles (percentile points marking 10% segments) of the 2008 TSS score distribution. We then used those same score points to categorize the 2012 and 2016 test takers. We calculated the relative percentages of the test takers falling into each of the categories and compared them to each of the 10% segments to determine where the largest differences were. Figure 5 illustrates the results.

Figure 5

Percentage of Applicants with Total Scale Scores Within Selected Ranges July Administrations



* The green line represents the passing score

As shown, over 21% of the 2016 test population is in the bottom decile of the 2008 TSS distribution (i.e., scores <=1290). The percentage rapidly decreases in the 2nd (1291-1353; 14%), 3rd (1354-1390; 13%) and 4th (1391-1445) deciles. In 2012, as scores were in the middle of their current decline, the percentages of the applicants in all four of these lower deciles were much more similar (ranging between 11% and 12%). Further, none of the other score ranges showed such wide differences between 2008, 2012 and 2016 as this lowest score range.

This finding leads to the question as to whether the composition of test takers at this lowest score level (i.e., <= 1290 and over 150 points from the passing standard) has systematically changed since 2016. To examine this question, we calculated the percentages of the applicants from various subgroups who fell into this group in 2016 and compared them to the percentages from the 2008 examination. Results are summarized in Table 10.

The entries in Table 10 represent the percentage of total test takers in the identified group that scored less than or equal to 1290. For example, in 2008, 6% of all students from CA-ABA schools had scores less than 1290 as compared 14% of all students from CA-ABA schools testing in 2016. The final column in the table presents the absolute differences in those percentages between the two years. In terms of total numbers, the 21% of total 2016 test takers in the lowest score band translates to almost 1,600 test takers.

Table 10

Percentage of Applicants in Various Subgroups with Total Scale Scores <= 1290 July Administrations

Subgroup	Year			
	2008	2012	2016	2008-2016
Examinees	N=858	N=1,045	N=1,578	
School				
CA ABA	6%	8%	14%	8%
Level I	7%	8%	18%	11%
Level II	3%	4%	9%	6%
Level III	2%	2%	7%	5%
Non CA ABA	8%	12%	18%	10%
CA Accredited	27%	32%	41%	14%
CA Non-Accredited	26%	35%	47%	21%
Foreign	42%	44%	57%	15%
Exams Taken				
1st	6%	8%	14%	8%
2nd	20%	24%	36%	16%
3rd	17%	18%	25%	8%
> 3rd	25%	31%	37%	12%
Racial/Ethnic				
Asian	11%	15%	27%	16%
Hispanic	16%	17%	24%	8%
Black	25%	26%	36%	11%
White	7%	9%	15%	8%
Gender				
Male	10%	12%	19%	9%
Female	10%	12%	21%	11%

These data shed additional light on changes in the composition of the test taking populations during the study period. In terms of the examinees' law schools, the relative percentage of students from CA-ABA in the lowest decile of the score range more than doubled from 6% in 2008 to 14% in 2016. Furthermore, it was the students from the Level I CA-ABA schools (lowest median LSAT) that accounted for the largest absolute change (11%). Statistics for NCA-ABA applicants mirrored those of the Level I CA-ABA applicants. Nearly 2 out of 5, and 1 out of 2 applicants from ACC and NAC schools scored in this lowest score range in 2016, compared to only 25% in 2008.

In terms of the impact of testing status, the percentage of 1st time takers in the lowest decile increased by 8% (more than doubling the rate) between 2008 and 2016. However, CBE first & third- time repeaters experienced the largest absolute increases , with about 1/3 of the 2016 applicants falling into the lowest score range.

The percentage of each racial/ethnic group falling into the <= 1290 range increased in 2016. By far, the largest change occurred among Asian students; roughly 10% had scored in the lowest decile in 2008, but almost three-times as many (27% or an absolute increase of 16%) did so in 2016. As a group, Blacks continued to have the largest proportion of applicants (36%) in the lowest score range while the relative increase was not as great as for Asians.

While the percentage of both males and females scoring in this group increased (doubling the percentage in 2008), the changes were roughly equivalent.

Results in this section lend evidence to the fact that decreases in CBE scores are not equivalent across the lower portions of the score distribution, and that the overall lower mean scores (and subsequent lower passage rates) may rather be a function of a large group of applicants sitting for the examination who are much less prepared, relative to applicants who took the CBE 9 years prior. The disproportionate increase in the percentage of applicants from selected subgroups (e.g., Level I ABA schools) who scored at these lower levels suggests that the 2016 applicant population may be substantively different (e.g., lower ability?) than those taking the 2008 exam.

4. Has the likelihood of <u>eventually</u> passing after 2 years changed over time?

The revised ABA accreditation process has proposed a standard requiring that 75% of a law school's graduating class pass the CBE within two years. Based on this standard and the fact that decreases in performance on the February exams (taken by disproportionately more repeaters), were not as great as in July examinations, we analyzed the available data to determine if the *changes* in the "two year" pass rates were as significant as the annual rate.

Our analyses tracked two cohorts of first time July CBE takers, one from 2008 and the other from 2014[13]. There were 6,235 and 6,185 first time applicants taking the July 2008 and 2014 CBE, respectively. Table 11 presents data on the outcomes for these two cohorts beginning with their initial attempt and 3 subsequent opportunities.

[13] The overall passing rate in 2014 was 49%, the first July examination that the rate dipped below 50% since the early 2000's.

Table 11

Bar Passage Rates after 2 Years
July 2008 vs. July 2014 First Time Test Takers

	% Pass on 1st Attempt	% Pass on Subsequent Attempt	% Pass Total	% Failing	% No Subsequent Attempts
2008	75%	11%	86%	8%	5%
2014	61%	19%	80%	12%	8%
2008-2014 Diff.	-14%	8%	-6%	4%	3%

Overall Eventual Pass Rates. Table 11 shows that for the 2008 cohort, 86% of the test takers passed the CBE within the 4 exam window; 75% on their initial attempt and an additional 11% on a subsequent attempt. Of the remaining applicants, 8% made one or more subsequent attempts and failed, while 5% did not make another attempt. For the 2014 cohort, 80% passed the CBE; 62% on their initial attempt and an additional 19% on their subsequent attempt. Of the remaining 2014 cohort, 12% failed on a subsequent attempt while 8% did not re-attempt testing.

Thus, while the difference for first time takers on their initial attempt was 14% between 2008 and 2014, the difference between the eventual passage rates after the four examination window was only 6%. Among those failing their first attempt, 5% did not reattempt in 2008 while 8% did not in 2014. Unfortunately the two-year passage rates for the July 2016 test takers will not be known for a few more years.

A. Eventual Pass Rates By Subgroups. Table12 shows the eventual pass rates by applicant subgroups.

Table 12

Subgroup Bar Passage Rates after 2 Years
July 2008 vs. July 2014 First Time Test Takers

Metric	2008	2014	2008-2014
Examinees	**N=858**	**N=1578**	
School			
CA ABA	94%	89%	-5%
Low LSAT	91%	81%	-10%
Medium LSAT	94%	90%	-4%
High LSAT	95%	93%	-2%
Non CA ABA	87%	78%	-9%
CA Accredited	55%	54%	-1%
CA Non-Accredited	45%	39%	-6%
Foreign	70%	63%	-7%
Racial/Ethnic			
Asian	85%	76%	-9%
Hispanic	80%	76%	-4%
Black	71%	65%	-6%
White	89%	85%	-4%
Gender			
Male	86%	81%	-5%
Female	86%	80%	-6%

Results from Table 12 suggest that after two years, the overall bar passage rates for the two cohorts converge, as do the rates within each of the subgroups. A difference of less than 10% in the two year success rates was observed for all of the subgroups in the two cohorts, and for several subgroups there was virtually no change. For example, there is only a 2% difference in the passage rates of applicants from Level III CA-ABA schools in the 2008 and 2014 cohorts (95% vs. 93%) and a 1% difference in the passage rate for students from ACC schools (55% vs. 54%). Historically lower performing groups (e.g., Foreign applicants, students from Level I schools, and some minority subgroups) tended to have slightly larger gaps in passage rates between the two time periods.

These findings indicate that there may be a decrease in the initial readiness of applicants or their preparation for taking the CBE since the 2008 examinations were given.

5. Have other statistical/psychometric properties of the examination changed over time in such a way that it may have impacted applicant scores?

Reliability measures the degree of stability or consistency of scores on a test and is one of any test's key psychometric property[14]. The lower the reliability, the higher the amount of error that exists in a measurement. Test reliability above .85 (out of 1.00) is considered acceptable for high stakes tests such as the CBE. Overall reliability on the CBE itself is a function of the separate reliabilities of the Written section, the MBE and the degree of correlation between the two. As any of these three values change, so does the reliability.

To determine whether there was any change in any of these metrics, we reviewed historical technical reports for the February and July CBEs in the study time frame. Table 13 summarizes data abstracted from these reports.

Table 13

Reliability Coefficients by Section and Total Test and Between Section Correlations For February and July CBEs

Year	July				February			
	Reliability			Correlation	Reliability			Correlation
	MBE	Written	Total	MBE & Written	MBE	Written	Total	MBE & Written
2008	.89	.80	.88	.68	.88	.75	.85	.55
2012	.90	.82	.88	.66	.89	.77	.86	.57
2016	.93	.82	.90	.73	.90	.78	.87	.61
2008-2016 Diff.	.04	.02	.02	.05	.02	.03	.02	.05

The overall Total Test reliability has remained quite high since 2008, increasing slightly (but not materially) in 2016. Overall reliability on the July administrations continues to slightly outpace February's, primarily due to the wider spread of scores on that administration. Increases in the overall reliabilities are a function of three factors. First, since 2008, the reliability of the MBE which has about ½ the weight (.35) as that of the Written section (.65) has steadily increased since 2008. Secondly, the reliability of the Written section has also increased slightly. And finally, the degree of relationship between the two sections increased over the same period (.68 to .73 on July CBEs and .55 to .61 on February CBEs), which is due in part to the increased reliability on the respective sections.

[14] Validity is another major psychometric property of a test. Data available to this study precludes an evaluation of any changes that may have occurred since 2008 in any of the various measures of validity that are used.

We can conclude from these findings that the consistency in scores, as measured by test reliability has not decreased over time, and has actually increased. The increasing correlations between sections on the exam would indicate that applicants are beginning to perform at more equivalent levels on the respective sections than in the past. This finding could dampen the compensatory nature of the current scoring method. However, none of these changes appear large enough to impact the decrease in scores and the subsequent passage rates.

6. How would the bar passage rates have changed if the cut point were set at standards used by other states?

Increasing concern voiced over California's high passing standard led us to ask how much the change in passing rates would have been impacted if California had adopted a lower passing score more in line with that used in other states. To conduct these analyses we focused on the July CBEs only. For each of the three years in the study timeframe we calculated the final TSS of all applicants and evaluated the distribution of those scores. We classified applicants as passing or failing using three different standards: 1) the current California standard of 144 (1440); 2) a standard of 135 (1350) which is used by the largest number of states in the country; and 3) a standard of 133 (1330), which is the standard currently used by New York. New York's standard was selected because the state tests the largest number of examinees in the country and is the only state testing more applicants than California. We then calculated the percentage of California applicants that would have passed under each of these standards for the July CBE in each of the three study years[15].

Results presented in Table 14 indicate that if the modal U.S. standard of 135 were used, 66% of all applicants would have passed the July 2016 CBE (i.e., 22% more examinees). This rate would be 15% lower than the estimated passing rate for the 2008 exam if the 135 standard was applied. Using a standard of 133, 7 out of 10 examinees would be estimated to pass and the difference from 2008 would drop to 13%.

The first- time passing rate provides a more direct comparison between the two time periods. At a 135 standard, 19% more first timers would have passed, and the difference between 2008 and 2016 would differ by only 13%. At a 133 standard, that difference is less than 10%.

Refining the comparison even further, we performed the calculations on first- time test takers from CA-ABA schools only (historically the best performing group of all applicants). Results are presented in Table 15.

[15] We acknowledge two limitations of these calculations. First, if alternative passing standards were used, different regrade bands may have been used. Second, some repeating applicants might have passed on an earlier attempt. We do not believe that the impact of these limitations is significant and that the directionality of results is valid.

Table 14

Actual and Estimated CBE Passage Rates
At Alternative Passing Points
July Examinations

Year	1st Time Taker			Repeater			All Examinees		
	144	135	133	144	135	133	144	135	133
2008	75%	89%	91%	28%	60%	68%	62%	81%	84%
2012	69%	86%	89%	18%	52%	61%	56%	77%	82%
2016	57%	76%	80%	17%	46%	54%	44%	66%	71%
2008-2016 Difference	18%	13%	9%	11%	14%	14%	18%	15%	13%

For test takers from Level I Schools, there remain large differences between 2008 and 2016 examinees (29%, 26% and 19% decreases at the three respective standards). However, the differences in the students from upper level schools paint a slightly different picture. At the modal standard (135) there is only a 9% difference in passage rates from 2008 to 2016 in Level II schools, and only a 7% difference in Level III schools. Over 85% of first time takers from these ABA schools would have passed on the July 2016 examination.

Table 15

Actual and Estimated CBE Passage Rates
At Alternative Passing Points
1st Time Takers at CA ABA Schools

Year	Level I Schools			Level II Schools			Level III Schools		
	144	135	133	144	135	133	144	135	133
2008	77%	93%	94%	83%	94%	96%	87%	96%	97%
2012	67%	88%	91%	76%	94%	95%	85%	95%	97%
2016	38%	67%	75%	64%	85%	89%	76%	89%	91%
2008-2016 Difference	29%	26%	19%	19%	9%	7%	11%	7%	6%

How would California applicants have fared relative to their counterparts in New York, all things held equal?[16] New York reports its general statistics (similar to California) after each administration (https://www.nybarexam.org/ExamStats/Estats.htm). From that site, we determined that closest type of comparison that could be made between California and New York was the bar passage rate of first time test takers from ABA approved institutions. We extracted these statistics for the same three July examinations included in our study time frame. We then calculated an estimated passing rate using the 133 standard that New York applies. The results are summarized in Table 16.

New York, which switched to the UBE in 2016 saw a 9% decrease (from 91% to 82%) in its passage rate between 2008 and 2016. For the CBE, when the 133 standard was applied to students who attended CA-ABA schools, fully 96% of those test takers would have passed the CBE in 2008, 95% in 2012 and 87% in 2016. The decrease between estimated 2008 and 2016 passage rates was 9%; identical to the New York drop. Further, within-year comparisons between the two states show California estimated to have passed 5% more candidates. It is interesting to note that several other states testing larger pools of applicants and having passing standards more closer to the modal mark of 135 (e.g., Texas, Massachusetts, Florida and New Jersey) all experienced decreases in their passing rates between 2008 and 2016 that ranged from about 8% to 12%.

Table 16

Actual New York & Estimated CBE Passage Rates
For 1st Time Test Takers
At ABA Schools

Year	New York Actual % Passing @ 133	California Estimated % Passing @ 133	Difference
2008	91%	96%	+5%
2012	85%	95%	+10%
2016	82%	87%	+5%
2008-2016 Difference	9%	9%	0%

[16] Note that in July 2016, New York switched to the Uniform Bar Examination (UBE) which included nationally administered written section along with the MBE. New York calculates its scale scores similarly to California's but it now gives its MBE and Written Section equal weighting.

V. SUMMARY AND CONCLUSIONS

A continuing drop in the percentage of applicants passing the California Bar Examination (CBE) has generated a considerable amount of public discussion. The trend has been nationwide and led to much debate about the underlying causes. Declining law school enrollments, changes in legal training curriculum, examination content and standards, and the quality and composition of examinees have all been cited as possible causes. Electronic CBE databases maintained by the Office of Admissions of the California Bar provided an efficient method of profiling where the declines have occurred as well as offering some initial insights into their causes.

Data on various characteristics of applicants and their CBE performance was abstracted from the existing databases for each of three years: 1) 2008, the year with the highest passage rate since 1997; 2) 2016 the most recent year for which data were available and when CBE results dropped to the lowest levels since before 1990; and 3) 2012, a midpoint between these two extremes. During the 9 year period there was an 11% decline in the number of July test takers and a corresponding 4% increase in February examinees, which historically include a higher proportion of applicants repeating the CBE than in the July administration. The relative mix of examinees also shifted between 2008 and 2016 as traditionally higher performing groups made up proportionately less of the total test takers. For the July administrations, first time test takers decreased by 6%, applicants from out of state ABA schools declined by 4%, and non-minority test-takers declined by 6%. Female test takers became the majority gender in 2016 as well.

Other key findings include the following:

- In terms of performance, the overall average Total Scale Scores (TSS) and bar passage rates dropped 66 points (1481 to 1415) points and 18% (62% to 44%) respectively for July applicants in 2016 as compared to 2008. The decrease was less pronounced for the February administration (13 points and 4%, respectively).

- The magnitude of the changes was not equal in all groups. For example, on the July CBE 1st time applicants passage rates dropped 18% versus 11% for repeaters; applicants from CA ABA schools with higher median LSAT scores dropped 11% as compared to an almost 30% decrease for applicants from lower LSAT schools.

- The drop in passage rates in the various racial/ethnic groups, however, varied by only 5%. Relatedly, the drop in scores on the Written and MBE sections were roughly equivalent within the various groups, suggesting that neither section disproportionately contributed to the change.

- Results from an estimation model indicated that all things being held equal, roughly 20% of the change in July CBE scores and 17% of the change in bar passage rates could be attributed to the change in the mix of test takers between 2008 and 2016.

Further exploration of the distribution of scores revealed that a highly disproportionate number of test takers scored at the very lowest levels of the distribution in 2016 relative to 2008 (21% vs 10%). A comparison of the composition of test takers scoring in lowest percentiles of applicants showed that while the percentage of all subgroups among these lowest performers increased between 2008 and 2016, there were relatively higher changes for some groups than others. For example, there was an 11% increase for the low LSAT school students compared to 5% from the higher level schools and 21% increase in Non-Accredited schools. Asians increased by 16% compared to half that in Hispanics and Whites.

To gain insights into applicant preparedness we examined bar passage *after two years,* reasoning that perhaps more recent candidates may not have been as prepared on their first attempt. A study of first time takers in July 2008 and 2014 showed that while the passage rates on the initial attempt for these years differed by a full 14% (75% vs 61%), the difference fell to 6% after a two year follow-up window (86% vs. 80%). The difference between two year pass rates (as compared to the one year rates) again tended to be relatively higher in historically lower performing groups.

Traditional psychometric characteristics of the test that could be measured with the available data showed no degradation in the Written, MBE or Total Test scores. Actually, the reliability coefficient increased slightly from 2008 to 2016, and the correlation in performance between different sections of the exam also rose (from .55 to .61) as a result. The magnitude of these changes would not have a material impact on passing rates.

Finally, the analysis of the impact of the passing standard (i.e., "cut score") on the 2008 to 2016 decrease revealed that the differences between the two years would have been projected to drop by 3% if the national modal standard (135) was used and 2% more if a standard of 133 was used. A direct comparison with New York (which is the only state that tests more applicants than California and also changed to the Uniform Bar Examination in 2016), using only 1st time ABA takers and the 133 standard, revealed identical 9% drops in the passing rates in both states. This finding lends supporting evidence refuting the contention that the decreases in passage rates were caused in part by California's non-adoption of the UBE.

These analyses suggest that while the change in composition of test takers and the passing standard itself may have led to some of the performance decreases between 2008 and 2016, there are most likely other factors in play. Institutional factors such as changes in curriculum and admission policies may have contributed. Also, completely unmeasured in this study are both latent legal ability of applicants and their law school performance. Our study used known correlates for these measures (often to limited groups of students) rather than individual student abilities.

From the available data, we cannot discern the degree to which these student-related factors have changed. However, some of the differences that were observed in this study between performances at the various levels of the CA ABA schools point to possible decreases. It is also possible that other qualitative factors such as poorer student study habits and decreased motivation may have played a role. Assessment of the nature, size and directionality of such factors require additional data.

This study also did not address whether the content of the CBE remains relevant to an assessment of minimum competency to practice law, or whether the current standard remains appropriate in today's practice environment. These are issues that require different data and different methods.

Due to the volume of public comments received, these have been posted online in three separate files:

- Public Comments Received via E-Mail:
  http://apps.calbar.ca.gov/cbe/docs/agendaItem/Public/agendaitem1000002007.pdf

- Public Comments Received via Online Comment Box:
  http://apps.calbar.ca.gov/cbe/docs/agendaItem/Public/agendaitem1000002008.pdf

- Public Comments Received from Other Sources:
  http://apps.calbar.ca.gov/cbe/docs/agendaItem/Public/agendaitem1000002009.pdf

Transcripts of the two days of public testimony are also posted online:

- August 14, 2017 Public Testimony:

  https://www.calbar.ca.gov/Portals/0/documents/communications/State-Bar-Public-Hearing-Transcript081417.pdf

- August 15, 2017, Public Testimony

  https://www.calbar.ca.gov/Portals/0/documents/communications/State-Bar-Public-Hearing-Transcript081517.pdf

		Simulated Cut Scores for July 2008 GBX					Simulated Cut Scores for July 2016 GBX				
		1330	**1350**	**1390**	**1414**	**1440**	**1330**	**1350**	**1390**	**1414**	**1440**
Total	# Passing	7,242	6,920	6,017	5,642	5,329	5,451	5,053	4,010	3,598	3,332
	% Passing	84.1%	80.4%	69.9%	65.5%	61.9%	70.9%	65.7%	52.1%	46.8%	43.3%
	% Increase*	*35.9%*	*29.9%*	*12.9%*	*5.9%*		*63.6%*	*51.7%*	*20.3%*	*8.0%*	
Gender											
Male	# Passing	3,795	3,617	3,121	2,911	2,756	2,679	2,484	1,970	1,760	1,635
	% Passing	83.8%	79.9%	68.9%	64.3%	60.9%	72.2%	66.9%	53.1%	47.4%	44.0%
	*% Increase**	*37.7%*	*31.2%*	*13.2%*	*5.6%*		*63.9%*	*51.9%*	*20.5%*	*7.6%*	
Female	# Passing	3,441	3,297	2,890	2,726	2,568	2,722	2,525	2,005	1,805	1,665
	% Passing	84.5%	80.9%	71.0%	66.9%	63.0%	69.5%	64.5%	51.2%	46.1%	42.5%
	*% Increase**	*34.0%*	*28.4%*	*12.5%*	*6.2%*		*63.5%*	*51.7%*	*20.4%*	*8.4%*	
Race/Ethnicity											
Asian	# Passing	1,520	1,435	1,205	1,113	1,046	1,161	1,066	835	735	676
	% Passing	81.8%	77.2%	64.8%	59.9%	56.3%	64.0%	58.8%	46.1%	40.5%	37.3%
	*% Increase**	*45.3%*	*37.2%*	*15.2%*	*6.4%*		*71.7%*	*57.7%*	*23.5%*	*8.7%*	
Black	# Passing	314	287	215	181	164	252	222	146	117	104
	% Passing	66.1%	60.4%	45.3%	38.1%	34.5%	49.8%	43.9%	28.9%	23.1%	20.6%
	*% Increase**	*91.5%*	*75.0%*	*31.1%*	*10.4%*		*142.3%*	*113.5%*	*40.4%*	*12.5%*	
Hispanic	# Passing	621	591	471	432	397	734	663	478	419	379
	% Passing	76.5%	72.8%	58.0%	53.2%	48.9%	65.7%	59.3%	42.8%	37.5%	33.9%
	*% Increase**	*56.4%*	*48.9%*	*18.6%*	*8.8%*		*93.7%*	*74.9%*	*26.1%*	*10.6%*	
White	# Passing	4,368	4,200	3,765	3,570	3,392	3,063	2,874	2,369	2,165	2,019
	% Passing	87.6%	84.3%	75.5%	71.6%	68.0%	77.7%	72.9%	60.1%	54.9%	51.2%
	*% Increase**	*28.8%*	*23.8%*	*11.0%*	*5.2%*		*51.7%*	*42.3%*	*17.3%*	*7.2%*	
Other	# Passing	98	91	71	67	60	100	93	66	56	52
	% Passing	79.0%	73.4%	57.3%	54.0%	48.4%	67.6%	62.8%	44.6%	37.8%	35.1%
	*% Increase**	*63.3%*	*51.7%*	*18.3%*	*11.7%*		*92.3%*	*78.8%*	*26.9%*	*7.7%*	

		Simulated Cut Scores for July 2008 GBX					Simulated Cut Scores for July 2016 GBX				
		1330	**1350**	**1390**	**1414**	**1440**	**1330**	**1350**	**1390**	**1414**	**1440**
First Time or Repeat											
First Time	# Passing	5,657	5,521	5,078	4,870	4,682	4,089	3,881	3,317	3,066	2,896
	% Passing	90.6%	88.4%	81.4%	78.0%	75.0%	79.5%	75.4%	64.5%	59.6%	56.3%
	*% Increase**	*20.8%*	*17.9%*	*8.5%*	*4.0%*		*41.2%*	*34.0%*	*14.5%*	*5.9%*	
Repeat	# Passing	1,585	1,399	939	772	647	1,362	1,172	693	532	436
	% Passing	67.0%	59.1%	39.7%	32.6%	27.3%	53.5%	46.0%	27.2%	20.9%	17.1%
	*% Increase**	*145.0%*	*116.2%*	*45.1%*	*19.3%*		*212.4%*	*168.8%*	*58.9%*	*22.0%*	
School Type											
ABA	# Passing	4,240	4,119	3,767	3,571	3,415	3,397	3,196	2,629	2,387	2,231
	% Passing	92.6%	90.0%	82.3%	78.0%	74.6%	82.5%	77.6%	63.8%	57.9%	54.2%
	*% Increase**	*24.2%*	*20.6%*	*10.3%*	*4.6%*		*52.3%*	*43.3%*	*17.8%*	*7.0%*	
CA Accredited	# Passing	458	408	265	225	196	356	294	169	131	100
	% Passing	61.5%	54.8%	35.6%	30.2%	26.3%	46.2%	38.1%	21.9%	17.0%	13.0%
	*% Increase**	*133.7%*	*108.2%*	*35.2%*	*14.8%*		*256.0%*	*194.0%*	*69.0%*	*31.0%*	
Registered	# Passing	194	165	107	88	76	111	95	44	38	35
	% Passing	60.8%	51.7%	33.5%	27.6%	23.8%	41.0%	35.1%	16.2%	14.0%	12.9%
	*% Increase**	*155.3%*	*117.1%*	*40.8%*	*15.8%*		*217.1%*	*171.4%*	*25.7%*	*8.6%*	
Out of State	# Passing	1,629	1,556	1,369	1,307	1,242	1,033	975	801	730	685
	% Passing	87.1%	83.2%	73.2%	69.9%	66.4%	72.9%	68.8%	56.5%	51.5%	48.3%
	*% Increase**	*31.2%*	*25.3%*	*10.2%*	*5.2%*		*50.8%*	*42.3%*	*16.9%*	*6.6%*	

* Percent increase of the number of applicants that would have passed under each simulated cut score level relative to the number of passing applicants under the curent cut score of 1440.

Summary Results of Five-Year Member Survey, 2017

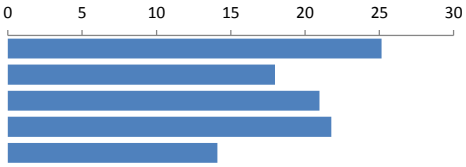
1. Response Rate and Time Taken to Complete the Survey

	# Delivered	# Responded	Response Rate	Average Time Taken (Min)
Full survey	112,899	8,562	7.6%	8.6
Short version 1 with reduced questions	56,489	4,356	7.7%	6.7
Short version 2 with reduced questions	56,499	4,343	7.7%	6.0
Total	225,887	17,261	7.6%	7.5

Attorney Background in Firm/Organization Type and Practice

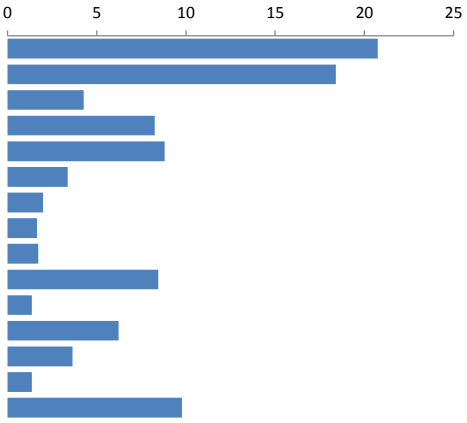
2. Years Since Admitted to the Bar

	%
1-10	25.2
11-20	18.0
21-30	21.0
31-40	21.8
41+	14.1



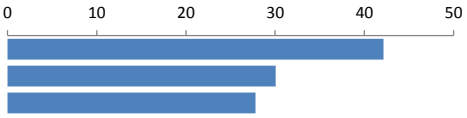
3. Which of the following best describes your current primary employment?

	%
Solo practice	20.8
Small firm	18.4
Medium firm	4.3
Large firm	8.3
Corporate in-house	8.8
Non-profit organization	3.4
District Attorney	2.0
Public Defender	1.7
Court	1.7
Other Government	8.5
Academic	1.4
Not employed as attorney	6.2
Other	3.6
Unemployed	1.4
Retired	9.8



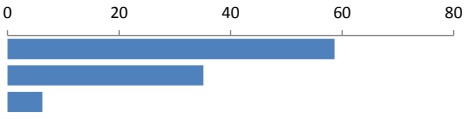
4. Are you an associate or partner?

	%
Partner	42.2
Associate	30.1
Other	27.8



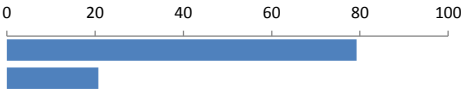
5. What type of legal entity is your private practice? (Sole practice and corporate in-house excluded)

	%
Limited partnership	58.6
Professional corporation	35.1
General partnership	6.2



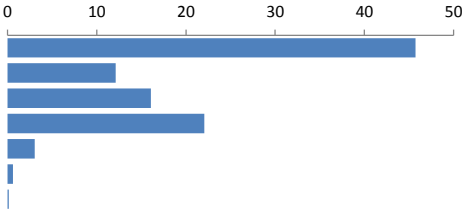
6. What type of legal entity is your private practice? (Sole practice only)

	%
Sole proprietorship	79.2
Professional corporation	20.8



7. What percentage of your clients are:

	%
Individuals	45.7
Small businesses	12.1
Medium businesses	16.1
Large businesses	22.1
Local government	3.0
State government	0.6
Federal government	0.1

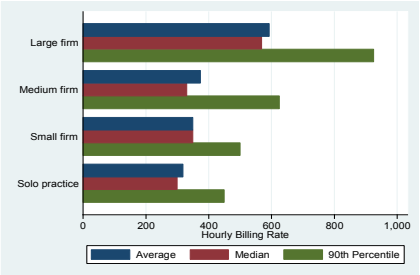


8. Client Type Served - by Firm/Org Type (%)

	Individuals	Small businesses	Medium businesses	Large businesses	Local government	State government	Federal government
Solo practice	71.1	14.5	8.6	3.8	1.1	0.5	0.2
Small firm	53.0	13.7	15.3	12.8	4.2	0.6	0.1
Medium firm	24.0	11.6	24.7	27.4	11.1	0.7	0.2
Large firm	8.4	8.8	26.3	51.3	3.6	1.2	0.1
Corporate in-house	5.0	4.9	23.9	65.3	0.3	0.4	0.2
Total	45.7	12.1	16.1	22.1	3.0	0.6	0.1

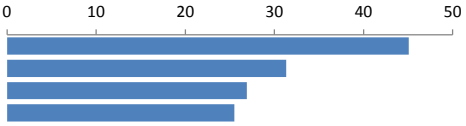
9. What is your average hourly billing rate?

	Mean	Median	90th %tile
Solo practice	318	300	450
Small firm	350	350	500
Medium firm	373	330	625
Large firm	592	568	925
Total	377	350	600

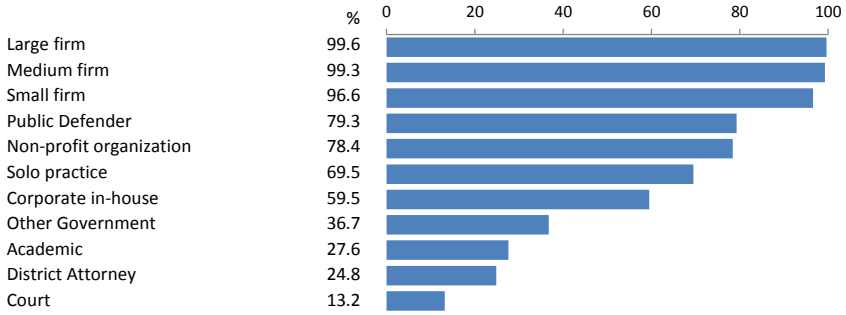


10. Do you provide services on an unbundled basis to clients, i.e., limited scope services--such as only drafting a motion or appearing at a hearing for an otherwise self-represented client? 35.7%

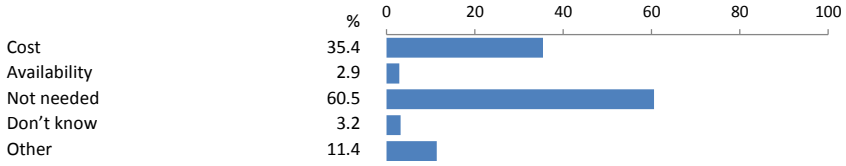
	%
Solo practice	45.1
Small firm	31.3
Medium firm	26.9
Large firm	25.5
Total	35.7



11. Are you covered by malpractice insurance? 66.2%

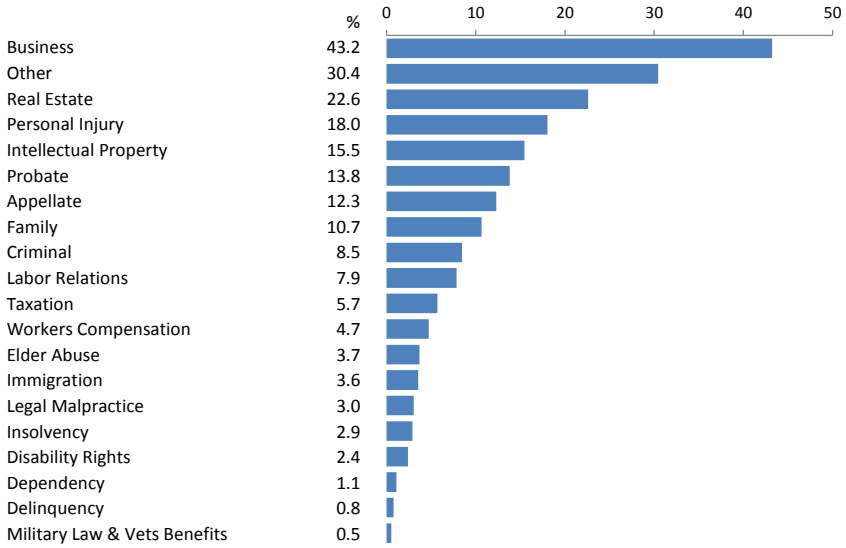


12. What are the reasons you are not covered by malpractice insurance?*



* Related to the response patterns in #11 above, cost was selected as a factor mostly by solo practitioners and those in small firms. More than 70 percent of "Not needed" responses came from corporate in-house attorneys or those in nonprofit or government agencies. The remaining 30 percent were selected by solo practitioners.

13. What are your areas of practice?



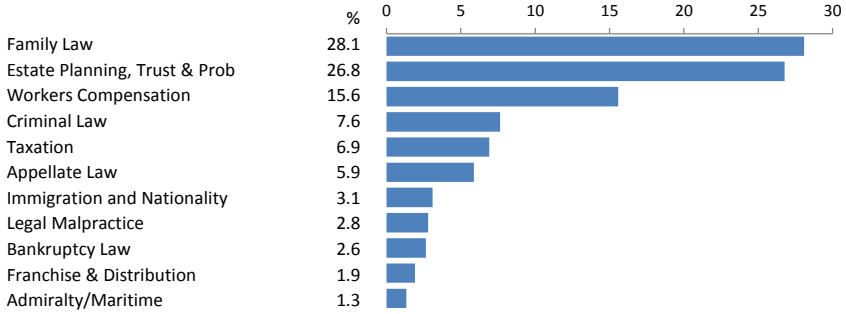
13a. Number of practice areas selected varies by firm type (%)

	Solo Practice	Small Firm	Midium Firm	Large Firm	Corporate In-house	Total
1	37.2	37.4	41.8	56.5	51.8	42.3
2	29.1	26.9	29.5	26.6	27.7	27.9
3	18.8	18.3	16.9	11.5	12.0	16.5
4 +	15.0	17.4	11.7	5.5	8.5	13.3
Total	100.0	100.0	100.0	100.0	100.0	100.0

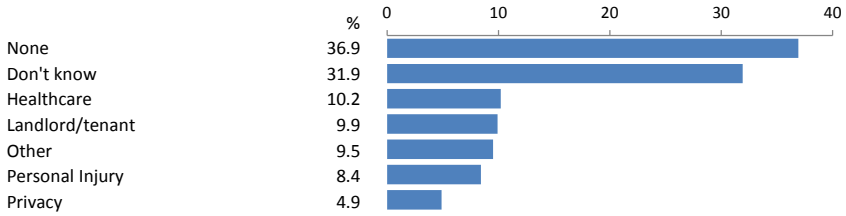
14. Are you certified by the State Bar in a legal specialty area? Percent Yes.* 4.4%

*More than 90 percent of those with a certified legal specialty came from the private sector, with approximately 40 percent each (of the total) from either solo practitioners or small firms.

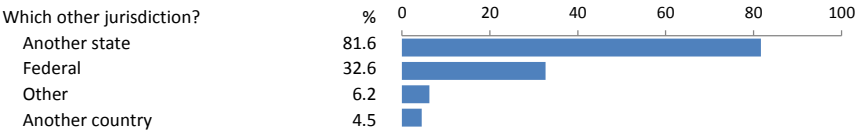
15. In which specialty areas are you currently certified?



16. Check any additional certified specialty areas that you think should be created.



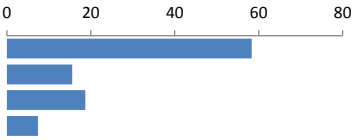
17. Are you licensed to practice law in another jurisdiction? Percent Yes. 24.3%



Pro Bono Legal Services

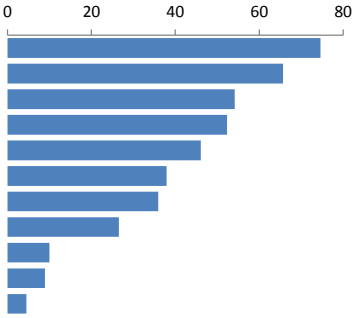
18. Do you provide pro bono legal services?

	%
No	58.4
Yes, to both individuals and nonprofit	15.5
Yes, directly to indigent individuals	18.7
Yes, to nonprofit organizations	7.4



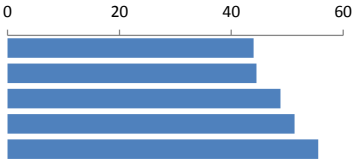
18a. Percent of respondents providing pro bono services varying by firm or organization type.

	%
Large firm	74.6
Solo practice	65.6
Small firm	54.2
Non-profit organization	52.4
Medium firm	46.1
Academic	37.9
Public Defender	35.9
Corporate in-house	26.5
Other Government	10.0
Court	8.9
District Attorney	4.5



18b. Varying by age as well as measured by years since admitted to the Bar.

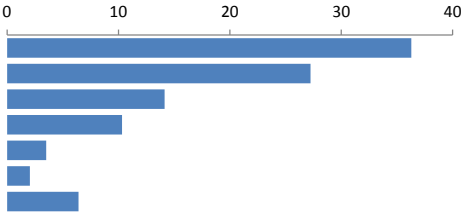
	%
1-10	44.0
11-20	44.5
21-30	48.8
31-40	51.3
41+	55.6



19. Approximately how many hours of pro bono legal work do you perform annually?

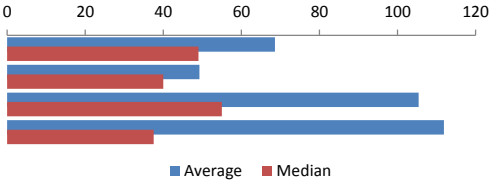
Average 70 hours and median 50 hours

	%
1-25	36.3
26-50	27.2
51-99	14.2
100-149	10.3
150-199	3.5
200-249	2.1
250+	6.4



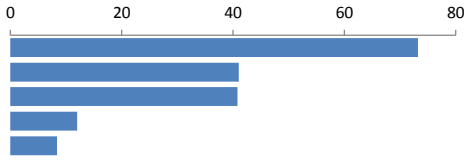
19a. With significant variance by employment background

	Average	Median
Private	69	49
Government	49	40
Nonprofit	105	55
Academic	112	38



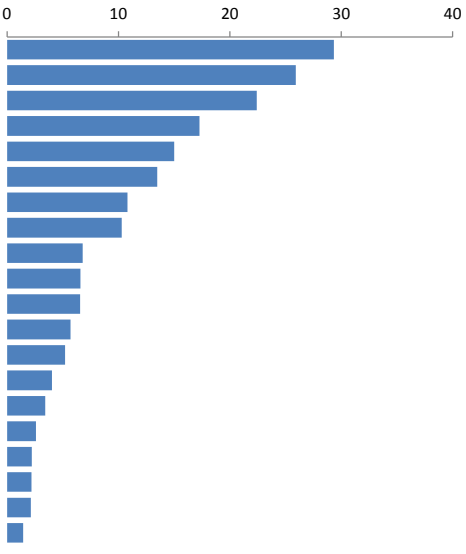
20. What types of pro bono services have you provided?

	%
Legal information & counsel	73.2
Full case representation	41.0
Limited scope representation	40.8
Co-counseling	12.0
Other	8.4

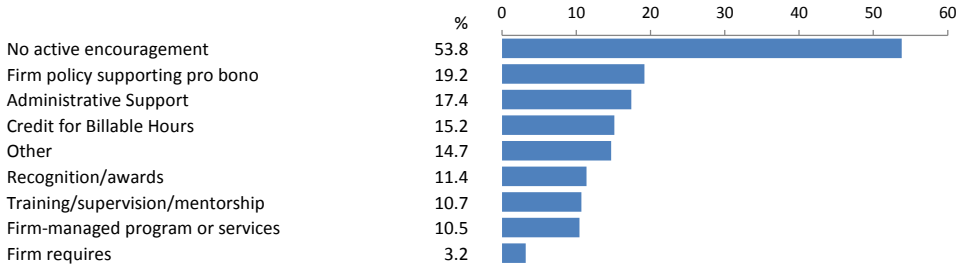


21. Please list the practice areas for your pro bono work.

	%
Business	29.3
Other	25.9
Family	22.4
Real Estate	17.3
Criminal	15.0
Immigration	13.5
Probate	10.8
Personal Injury	10.3
Labor Relations	6.8
Disability Rights	6.6
Intellectual Property	6.6
Elder Abuse	5.7
Taxation	5.2
Insolvency	4.0
Military Law & Vets Benefits	3.4
Appellate	2.6
Delinquency	2.2
Dependency	2.2
Workers Compensation	2.1
Legal Malpractice	1.5



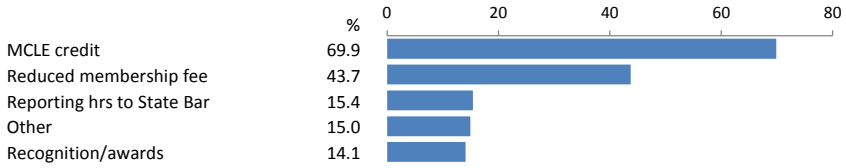
22. What does your employer do to encourage you to perform pro bono work?



22a. Different incentives for pro bono services varying across employment sectors.

	Private	Gov't	Nonprofit	Academic
No active encouragement	47.5	75.5	43.1	58.5
Firm policy supporting pro bono	29.5	2.9	9.8	3.5
Administrative Support	26.4	2.5	8.9	11.3
Credit for Billable Hours	23.8	1.3	4.3	1.4
Other	6.8	16.7	33.3	20.4
Recognition/awards	16.5	3.1	6.4	12.7
Training/supervision/mentorship	15.5	2.3	10.1	3.5
Firm-managed program or services	16.0	1.3	5.5	0.0
Firm requires	5.1	0.1	1.5	0.0

23. Is there anything the State Bar can do to help you provide pro bono legal services?

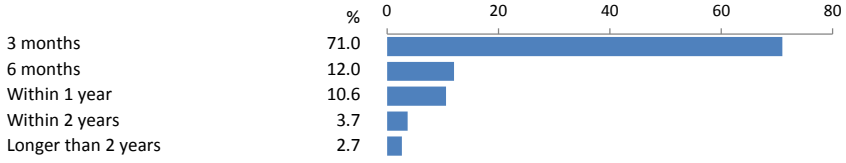


23a. Views regarding what the Bar can do in support of pro bono services varying across employment sector.

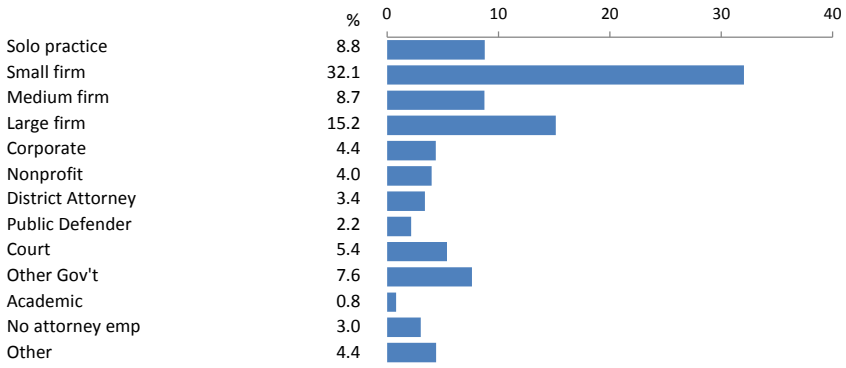
	Private	Gov't	Nonprofit	Academic
MCLE credit	76.7	55.5	74.1	54.7
Reduced membership fee	41.7	44.2	53.8	55.7
Reporting hrs to State Bar	16.2	13.6	25.9	19.8
Other	10.3	23.8	12.6	15.1
Recognition/awards	14.6	17.0	24.1	21.7

Career Path Since Law School

24. How soon after graduation from law school were you able to obtain law-related paid employment?



25. What type of paid employment did you first obtain after graduation from law school?



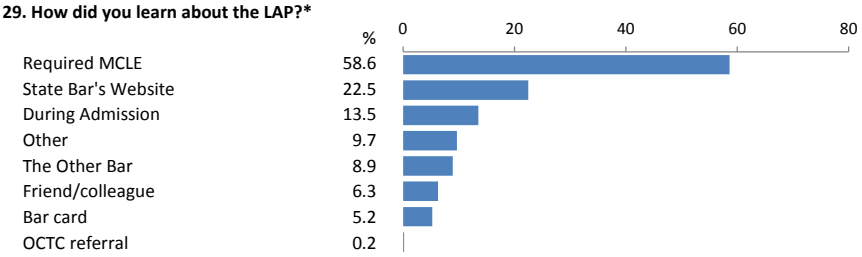
26. Since graduating from law school, approximately how many different law-related paid jobs have you had? Average (and median) 3

27. How long did you typically stay in each job (years)? Average 9.7 years and median 5 years

Awareness and Views Regarding Lawyer Assistance Program (LAP) and Minimum Continuing Legal Education (MCLE)

28. Are you aware of the confidential services offered by the Lawyer Assistance Program (LAP) to lawyers suffering from substance abuse or mental health issues? Percent Yes: 72%

29. How did you learn about the LAP?*

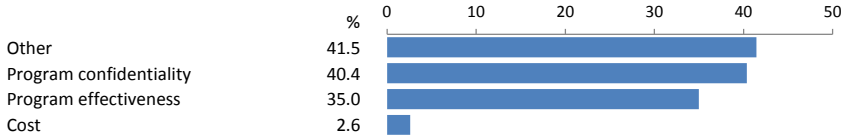


* The most common comments provided for the "Other" answer referred to law school as the source or considered it general knowledge. "Do not recall" is also a common response.

30. Have you ever used the services of the LAP? Percent Yes. 1.9%

31. If you had a friend or colleague in the legal profession who you thought was struggling with substance abuse or mental health problems, would you refer them to the LAP? Percent Yes. 85.1%

32. Why wouldn't you refer a friend or colleague to the LAP?**



** A large number of those who selected "other" as the reason noted the existence of other programs that they thought might be more suitable; many also thought that it was ultimately the decision of the individual who needs it.

32a. Both gender and age (years since admitted to the Bar) play a role in the different weights given to confidentiality and

	Confidentiality	Effectiveness
Male	37.6	34.7
Female	31.4	44.8

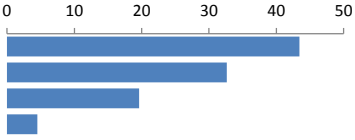
32b. Years since admitted to the Bar

1-10	38.6	46.2
11-20	33.0	42.4
21-30	30.0	43.3
31-40	35.4	39.1
41+	47.5	30.3

33. Have you ever sought assistance for personal concerns about substance abuse or mental health? 13.4%

34. What percentage of your Minimum Continuing Legal Education (MCLE) is provided through:

	%
Live class training	43.4
Online	32.6
Employer (either live or online)	19.6
Other	4.5

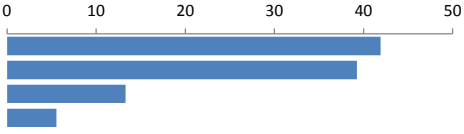


34a. With significant variance across employment types

	Employer	Online	Live Class	Other	Total
Solo practice	2.3	43.8	49.2	4.6	100
Small firm	7.8	34.7	54.3	3.3	100
Medium firm	31.1	23.4	42.2	3.5	100
Large firm	51.2	19.3	27.5	2.3	100
Corporate in-house	18.7	37.2	42.7	2.0	101
Non-profit organization	20.5	31.0	45.8	2.5	100
District Attorney	68.6	6.0	25.1	1.3	101
Public Defender	67.3	6.6	24.9	1.0	100
Court	38.2	18.2	35.5	7.9	100
Other Government	40.7	17.0	30.9	11.8	100
Academic	13.4	31.6	27.5	27.4	100

35. Who pays the costs of your MCLE courses?

	%
I do	41.9
My firm/organization	39.3
Both	13.3
Only free MCLE courses	5.5

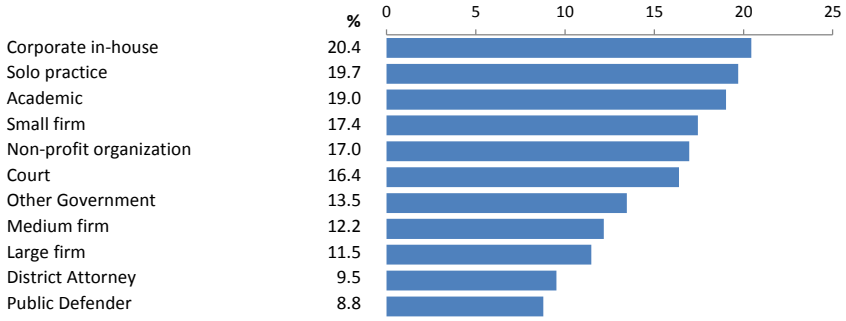


35a. With significant variance across employment types as well

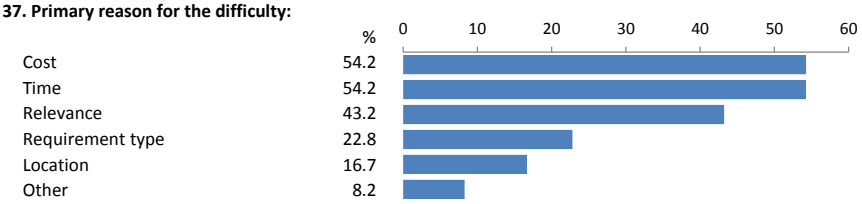
	I do	My Firm	Both	Only free MCLE courses
Solo practice	88.3	6.5	3.8	1.4
Small firm	23.4	59.5	16.6	0.5
Medium firm	7.8	73.5	17.4	1.4
Large firm	4.0	78.3	15.9	1.9
Corporate in-house	12.9	64.8	15.7	6.6
Non-profit organization	18.8	48.6	21.9	10.8
District Attorney	3.1	67.5	18.9	10.5
Public Defender	12.7	44.8	38.7	3.9
Court	23.0	37.9	27.0	12.1
Other Government	17.3	46.3	20.9	15.5
Academic	47.2	15.8	20.5	16.5

36. Do you find it difficult to comply with the requirement to complete 25 hours of MCLE every 3 years? Percent Yes. 19.2%

36a. With noticeable difference across employment types



37. Primary reason for the difficulty:

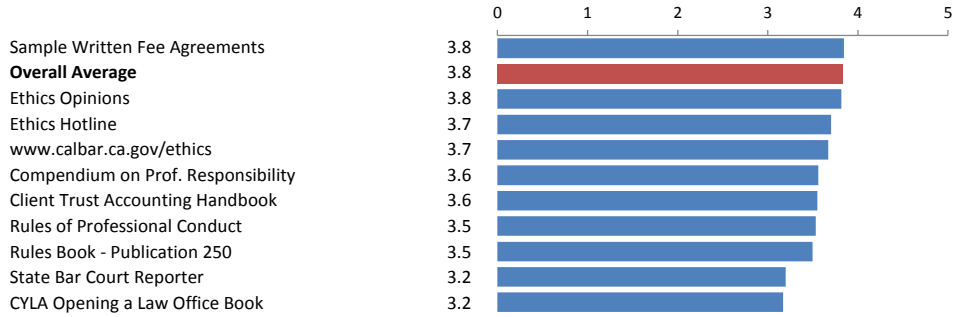


37a. Reasons cited varying across employment types

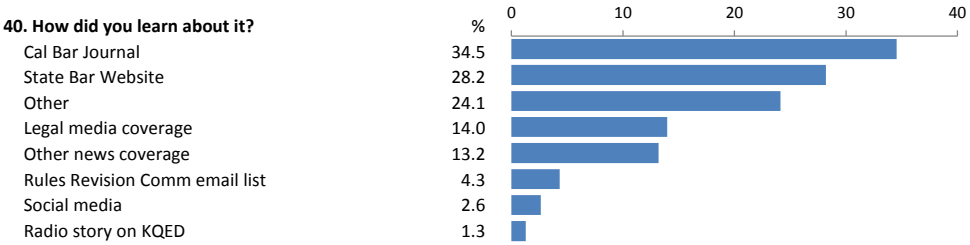
	Cost	Time	Relevance	Requirement Type	Location	Other
Solo practice	65.2	54.3	46.7	22.7	16.0	8.0
Small firm	43.0	65.9	41.1	20.6	18.0	4.9
Medium firm	41.0	73.8	41.0	29.5	6.6	3.3
Large firm	20.0	66.4	40.9	37.3	11.8	6.4
Corporate in-house	34.4	63.7	47.6	25.9	15.6	7.5
Non-profit organization	65.2	43.9	56.1	37.9	7.6	10.6
District Attorney	31.8	81.8	27.3	36.4	9.1	9.1
Public Defender	68.8	56.3	25.0	31.3	25.0	6.3
Court	60.0	56.7	36.7	26.7	16.7	6.7
Other Government	58.9	51.8	36.6	29.5	17.0	12.5
Academic	63.0	33.3	63.0	22.2	33.3	7.4

Satisfaction with Bar Resources for Improving Professional Competence

38. Please rate your level of satisfaction for any of the following State Bar professional responsibility resources that you have used. (1-5 scale)



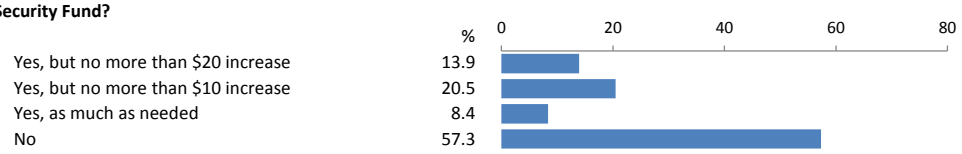
39. Are you aware that the State Bar is conducting a study of the Rules of Professional Conduct with the goal of submitting a recommendation for comprehensive rule amendments to the Supreme Court of California by March 31, 2017? Percent Yes. 19.6%



Experience with the Client Security Fund and the Attorney Discipline System

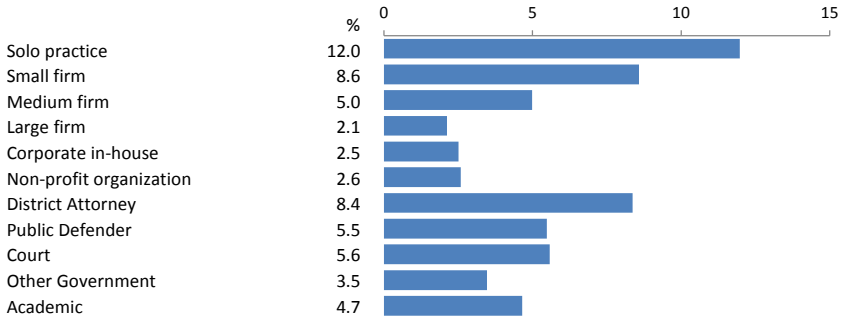
41. Have you ever advised someone who had money misappropriated by an attorney to file an application with the Client Security Fund? Percent Yes. 7.7%

42. Would you support an increase to the annual assessment that active attorneys pay to provide additional funds to the Client Security Fund?

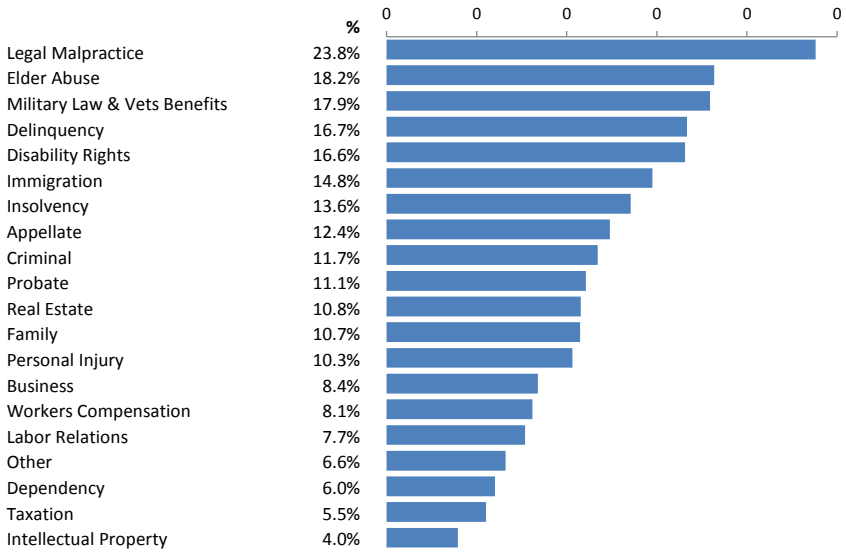


43. Have you ever filed a complaint with the State Bar against another attorney? Percent Yes. 6.9%

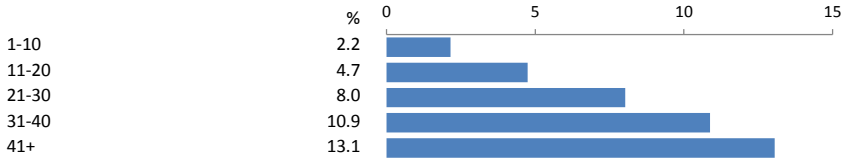
43a. With solo practice attorneys having the highest chance of having filed a complaint



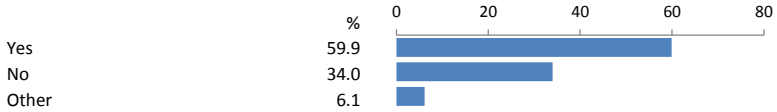
43b. The probability of filing a complaint varying across practice areas



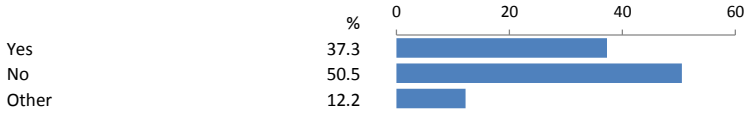
43c. Also showing difference across age groups as measured by the number of years since admitted to the Bar



44. Were you informed of the outcome/resolution of your complaint(s)?

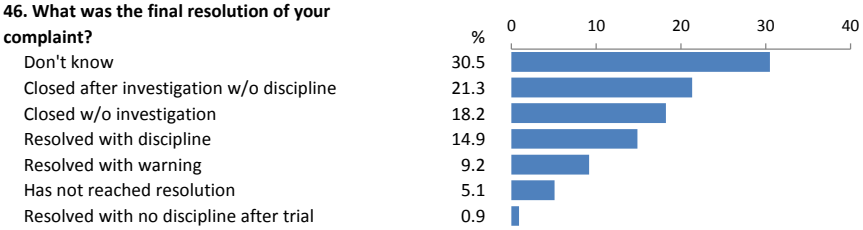


45. Positive Opinion of the Process?*



*No difference was found across either employment firm type or age groups as measured by years since admitted to the Bar.

46. What was the final resolution of your complaint?

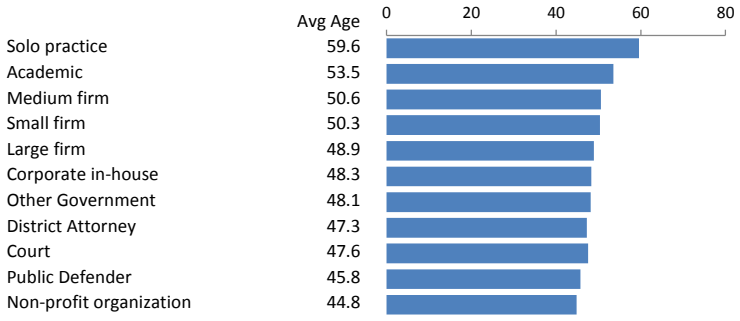


47. Did you attempt to address the misconduct you observed in another way? Percent Yes. 52.9%

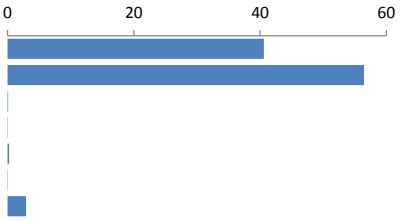
Demographics

48. Age of Respondents: average 54, median 55

48a. With significant difference across employment types



49. Gender Identify	%
Female	40.6
Male	56.5
Transgender female	0.1
Transgender male	0.1
Gender variant	0.2
Not listed	0.1
Prefer not to answer	3.0

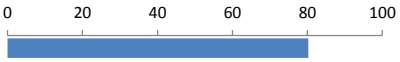


49a. Changing Composition of Gender Identity Groups Over Time (% within age group)

	Number of years Admitted to the Bar					
	1-10	11-20	21-30	31-40	41+	Total
Female	52.7	49.6	44.0	34.3	12.4	40.7
Male	44.7	46.9	52.7	62.4	85.9	56.3
Transgender female	0.1	0.1	0.1	0.0	0.1	0.1
Transgender male	0.1	0.0	0.0	0.1	0.0	0.1
Gender variant	0.6	0.3	0.1	0.1	0.0	0.2
Not listed	0.2	0.0	0.0	0.1	0.0	0.1
Prefer not to answer	2.5	3.1	3.7	3.3	1.9	3.0
Total	100.9	100.2	100.6	100.3	100.2	100.5

50. Sexual Orientation	%
Heterosexual	86.4
Gay/lesbian	4.4
Bisexual	1.3
Not listed	0.2
Prefer not to answer	7.5
More than one	0.3

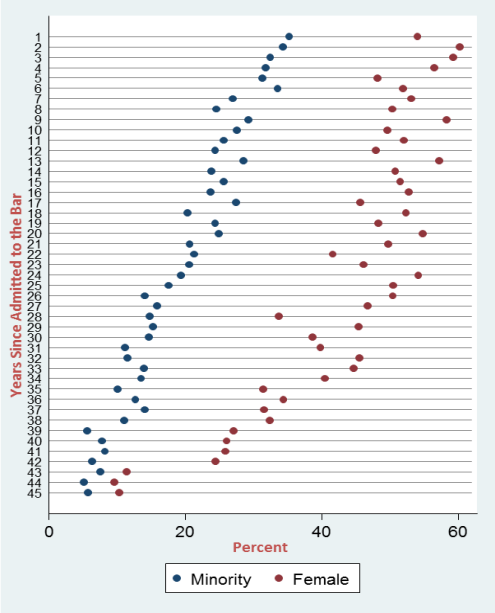
51. Race/Ethnicity Background	%
White	80.3
Asian	5.9
Hispanic	4.6
More than one	3.4
Other	3.3
Black	1.9
Native Hawaiian	0.3
Am Indian	0.3



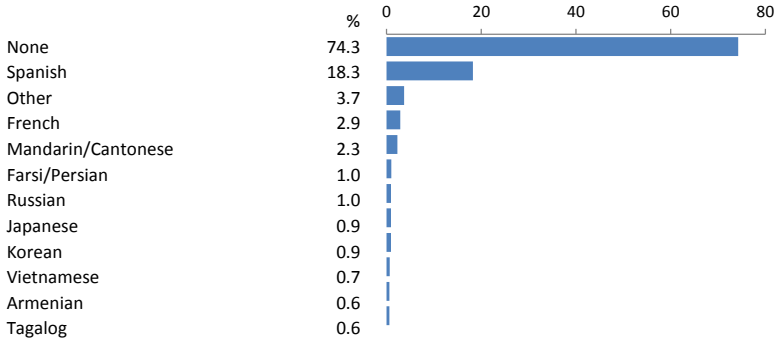
51a. Changing composition of race/ethnicity over time (% within age group)

	Number of years Admitted to the Bar					
	1-10	11-20	21-30	31-40	41+	Total
Am Indian	0.3	0.4	0.2	0.3	0.3	0.3
Asian	10.9	8.4	4.3	2.2	1.1	5.8
Black	2.2	2.3	2.6	1.0	0.7	1.8
Hispanic	6.8	5.7	4.5	2.6	1.4	4.5
Native Hawaiian	0.3	0.7	0.3	0.2	0.1	0.3
White	69.1	75.1	82.8	88.8	94.0	80.8
Other	3.2	3.9	3.4	3.2	1.7	3.2
More than one	7.1	3.5	2.1	1.8	0.7	3.4
Total	100.0	100.0	100.0	100.0	100.0	100.0

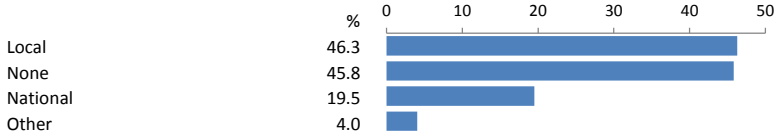
51b. Diversity of the legal profession in gender and race/ethnicity has been growing steadily over time.



52. What languages other than English do you use in your legal practice?



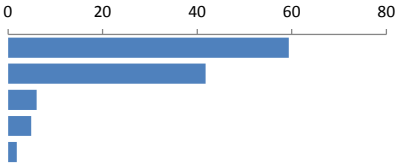
53. Do you belong to any voluntary Bar Association?



54. I identify as a person with a disability. Percent Yes. 4.5%

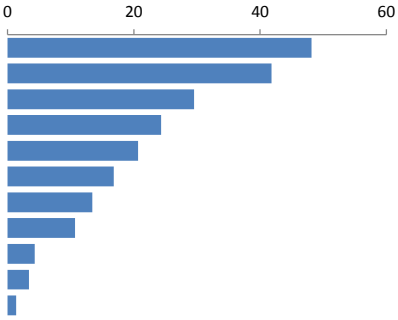
55. With a significant disability:

	%
Deaf	59.4
Blind	41.8
W/o use of hands or limbs	6.0
Difficulty walking	4.9
Other	1.9



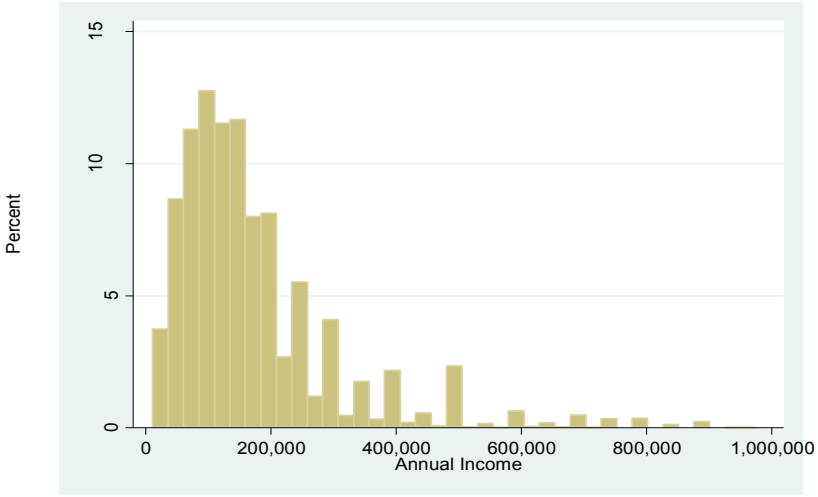
56. How can these disabilities be accommodated in your practice?

	%
Part-time	48.2
Permitting telework	41.8
Modifying physical space	29.5
Modifying workplace policies	24.3
Other	20.7
Job restructuring	16.8
Unpaid leave	13.4
Changing methods of supervisio	10.7
Materials in alternative format	4.3
Reassignment	3.4
Qualified readers or sign langua	1.4



57. Are you a veteran who served in the active military, naval, or air service, and who was honorably discharged? 9.8%

58. Approximate annual income received from legal practice: average $193,000, median $135,000



PR-11-04

THE ESTIMATED EFFECT ON EXAMINATION QUALITY AND PASSING RATES OF DIFFERENT WAYS OF MODIFYING CALIFORNIA'S BAR EXAMINATION

Stephen P. Klein, Ph.D. & Roger Bolus, Ph.D.
December 12, 2011

Overview

California's General Bar Examination (GBX) is an 18-hour (three-day) test. It consists of the MBE (which is a 6-hour 200-item multiple choice test), a set of six 1-hour essay questions, and two 3-hour Performance Test (PT) questions. This report estimates the likely effects on this exam's quality and passing rate if it was shortened to a two-day test and gave the MBE and written sections equal weight.

Samples

The population for our analyses consisted of all the applicants who took the GBX one or more times between 2001 and 2010. There were 43,832 February and 81,346 applicants in this 20-exam sample for a grand total of over 125,000 applicants. We also analyzed the essay and PT scores of the subset of 20 to 25 applicants who had their answers to each question graded independently by at least ten of the readers assigned to a question; i.e., all the applicants in this sample had their answers to each question graded ten times.[1]

Purposes & Definitions

Our analyses examined how score reliability was affected by: (a) using two or more independent readers per answer, (b) giving the MBE 50% of the weight (instead of the current 35%) in determining an applicant's total exam score, and (c) shortening the written portion of the exam from a two-day 12 hour test to a one day six or seven-hour test. We also examined the percentage of applicants in different racial/ethnic and gender groups whose pass/fail status would be affected by the number and sample of essay and PT questions they answered.

The term "score reliability" in this study refers to the likelihood that applicants would receive the same *score* (as distinct from pass/fail decision) regardless of the particular set of California bar exam essay and PT questions they were asked or the set of readers who graded their answers. For example, an essay test with high score reliability is one where the applicants who earn relatively high scores on one question also tend to earn relatively high scores on the test's other questions. All other things being equal, the higher the score reliability the greater the confidence that can be placed in the results.

[1] Results with this sample and those who went to reread must be treated very cautiously because they are not random or representative samples of the population of all takers.

"Score reliability" (coefficient alpha) is reported on a 0.00 to 1.00 scale with 1.00 being best. Values less than 1.00 may be due to: (a) some applicants being more proficient in the skills and knowledge needed for some questions while other applicants have a different pattern of expertise, (b) differences among readers in the score they would assign to an answer, and (c) other factors, such as how much scores spread out from the mean. In California, adjusting for the typically small difference in means and standard deviations among readers on a question usually has little or no effect on the written test's score reliability.

"Decision consistency" (which is an especially important characteristic of a licensing test) refers to the stability of pass/fail decisions, such as across different types of tests or versions of a test. Thus, it is a useful index for examining the comparability of different test designs. Decision consistency is highest when: (a) score reliability is high and (b) the passing rate is well above or well below 50%.

Effect of Number of Readings Per Answer on Score Reliability

Score reliability increases as the number of readers per answer increases, but the benefit of additional readers tapers off rapidly. For example, the first row of Table 1 shows that the reliability coefficient for a 6-question essay test in July is 0.06 points higher with two readers than it is with one reader, but adding a third reader results in only a 0.01 improvement over having two readers. In short, the marginal benefit of adding readers disappears quickly (although it seems to be greater for two 3-hour PT questions than for a set of six 1-hour essay questions).

Table 1
Increase in Score Reliability Over a Single Reader as a Function of the Number of Additional Readers per Answer, Type of Question Asked, and Test Month

Number of additional readers	6 Essay Questions			2 PT Tasks	
	February	July		February	July
1	.06	.06		.06	.07
2	.07	.07		.11	.17
3	.07	.08		.13	.17
4	.07	.08		.15	.17
5	.08	.08		.17	.17
6	.08	.08		.18	.17
7	.09	.09		.18	.17
8	.09	.09		.18	.17
9	.10	.09		.18	.17

Applicants have three hours per PT task and an average of one hour per essay question. Results in this table are based on the answers written by the 20 to 25 applicants who had all of their answers graded by all the readers assigned to each question.

The diminishing benefit to improving score reliability by having more than two readers per answer supports California's policy of having a second reading of all the answers written by all the candidates who came close to passing but failed after the initial reading of their answers. In addition, the limited benefit of additional readings suggests that the less than perfect reliability of the written test stems mainly from an interaction between applicants and questions rather than from differences among readers in the scores they would assign to an answer.

The remaining analyses in this report are based on just the first reading of an applicant's answers in the ten-year population of February and July takers. We did this because: (a) not all applicants had their answers read at least twice and (b) which applicants would have their answers read more than once was likely to vary across the different test designs we examined. Thus, the results with these models may underestimate the score reliability that is likely to occur if California continues rereading the answers of those who initially came close to passing.

Reliability of Essay and PT Scores

The mean reliability of a single reading of a set of six 1-hour essay questions in our population of February and July takers was 0.64 and 0.70, respectively. The higher scorer reliability in July than in February may be due at least in part to the greater variance in July scores. In both months, the reliability of the sum of the scores on a single reading of two 3-hour PTs was about 0.52 (based on the Spearman-Brown stepped-up mean correlation of 0.35 between two PT scores).

Procedures for Computing Total Scores

MBE raw scores (i.e., the number of items answered correctly) are converted to scale scores to adjust for possible differences in the difficulty of the questions asked. Essay and PT readers assign scores to answers in 5-point intervals on a 40 to 100-point scale. PT raw scores are then multiplied by 2.00 so that the maximum possible written raw score is 1,000 points. California (like most other states) converts its written raw scores to a score distribution that has the same mean and standard deviation as its MBE scale scores. This step adjusts the reader assigned total raw scores for possible variation in essay and PT question difficulty and grading standards over time. Total scale scores are computed using the formula below. Applicants with total scores of 1440 or higher pass, those in the 1390-1439 zone have all their answers reread, and all others fail.

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

Except as noted otherwise, the same procedures were used to compute written scale and total scale scores and to determine an applicant's pass/fail status for all the models discussed in the next section of this report.

Modeling Results

Tables 2 and 3 should be used together. Table 2 lists the key features of the models we examined and Table 3 shows their impact on total (MBE + Written) score reliability on February, July, and all exams combined.[2] For example, the only difference between models 1a and 1b is that as per current practice, model 1a weights the written and MBE scores 65% and 35%, respectively. In contrast, model 1b weights them equally. Table 3 shows that this single difference results in a relatively large improvement in score reliability (0.06 in February and 0.05 in July). The benefits of going to 50/50 weighting are consistent with the differences in reliability between models 5a and 5b.

Models 2a and 2b have the same structure, namely: three 1-hour essay questions and one 3-hour PT with the MBE and written sections weighted equally. The only difference between these models is that they use completely different essay and PT questions. The degree of agreement between these models therefore provides an unbiased estimate of their decision consistency and shows the reliability of an exam that is limited to the MBE and a 6-hour written test composed of three 1-hour essay questions and one full 3-hour PT question when the MBE and written portions are weighted equally.

Models 4a and 4b show the results for a two-day exam consisting of five essay questions and one PT. Although these analyses had to rely on data from 3-hour PTs, the results with them are likely to be very close to what would be obtained with 90-minute PTs; i.e., a 6½ hour written test. Models 4a and 4b have higher reliabilities than the current exam (model 1a) as a result of their giving the MBE and written sections equal weight.

Tables 4 and 5 show pass/fail decisions are consistent between various pairs of models. For example, Table 5 shows that in July, 93% of the applicants had the same pass/fail status under Model 2 (a two-day exam with a written component consisting of 3 essay questions and one PT) as they had with the current exam (i.e., a test with twice as many essay and PT questions) provided both exams weighted the written and MBE sections equally.

Table 6 shows that reducing test length does not affect overall passing rates or exacerbate the differences in rates that are typically found among racial/ethnic groups. Assigning equal weights eliminates the difference in passing rates between men and women. In short, *California can implement a two day exam in a way that improves test quality, maintains existing pass/fail standards, and does so without making it more difficult for minority applicants to pass*.

[2] Total score reliability calculations used MBE score reliabilities of .89 and .91 for the February and July exams, respectively as per the mean values in the MBE's technical reports. Written test reliabilities (coefficient alphas) were based on un-standardized essay raw scores on the first reading.

Table 2
Main Features of the Models Tested

Model	Essay	PT	Written/MBE Weights	Written Time	Model Description
1a	1-6	A & B	65/35	12 hrs	Current model & 65/35 weights
1b	1-6	A & B	50/50	12 hrs	Current but 50/50 weights
2a	1-3	A	50/50	6 hrs	Half of current written exam
2b	4-6	B	50/50	6 hrs	Half of current written exam
3a	1-4	A	50/50	7 hrs	4 1-hr Essays + one 3-hr PT
3b	3-6	B	50/50	7 hrs	4 1-hr Essays + one 3-hr PT
4a	1-5	A	50/50	8 hrs	5 1-hr Essays + one 3-hr PT
4b	2-6	B	50/50	8 hrs	5 1-hr Essays + one 3-hr PT
5a	1-6	None	65/35	6 hrs	6 1-hr Essays 65/35 weights
5b	1-6	None	50/50	6 hrs	6 1-hr Essays 50/50 weights
6	None	A&B	50/50	6 hrs	PT only

Table 3
Total Score Reliability (Coefficient Alpha, decimal points omitted)

Model Number	Test Month(s)			Model Description
	February	July	All	
1a	81	85	83	Current model & 65/35 weights
1b	87	90	88	Current but 50/50 weights
2a	80	85	82	Half of current written
2b	80	83	82	Half of current written
3a	82	87	84	4 1-hr Essays + one 3-hr PT
3b	82	86	84	4 1-hr Essays + one 3-hr PT
4a	84	88	86	5 1-hr Essays + one 3-hr PT
4b	84	88	86	5 1-hr Essays + one 3-hr PT
5a	81	85	83	6 1-hr Essays 65/35 weights
5b	86	89	88	6 1-hr Essays 50/50 weights
6	78	80	79	PT only

Table 4
Average Percentage of **FEBRUARY** Applicants with the Same Pass/Fail Status Under Alternative Models

Model 1a	Model 1b	% Agree
MBE weighted 35% Essays 1-6 in 6 hours PT-A & B in 6 hours Reliability = .81	MBE weighted 50% Essays 1-6 in 6 hours PT-A & B in 6 hours Reliability = .87	95%
Shows unique effect of weighting the MBE 50%		

Model 1b	Mean of Models 2a & 2b	% Agree
MBE weighted 50% Essays 1-6 in 6 hours PT-A & B in 6 hours Reliability = .87	MBE weighted 50% 3 Essays in 3 hours 1 PT in 3 hours Reliability = .80	91%
Models 2a and 2b cut test length in half with MBE weighted 50%		

Model 2a	Model 2b	% Agree
MBE weighted 50% Essays 1-3 in 3 hours PT-A in 3 hours Reliability = .80	MBE weighted 50% Essays 4-6 in 3 hours PT-B in 3 hours Reliability = .80	82%
Same models but completely different written questions in 6 hrs		

Model 3a	Model 3b	% Agree
MBE weighted 50% Essays 1-4 in 4 hours PT-A in 3 hours Reliability = .82	MBE weighted 50% Essays 3-6 in 4 hours PT-B in 3 hours Reliability = .82	86%
Models share 2 of their 4 essay questions in 7 hrs		

Model 4a	Model 4b	% Agree
MBE weighted 50% Essays 1-5 in 5 hours PT-A in 3 hours Reliability = .84	MBE weighted 50% Essays 2-6 in 5 hours PT-B in 3 hours Reliability = .84	88%
Proxy for a 6½ hour written exam (4 essay questions in common)		

Table 5
Average Percentage of **JULY** Applicants with the
Same Pass/Fail Status Under Alternative Models

Model 1a	Model 1b	% Agree
MBE weighted 35% Essays 1-6 in 6 hours PT-A & B in 6 hours Reliability = .85	MBE weighted 50% Essays 1-6 in 6 hours PT-A & B in 6 hours Reliability = .90	96%
Unique effect of weighting the MBE 50%		

Model 1b	Mean of Models 2a & 2b	% Agree
MBE weighted 50% Essays 1-6 in 6 hours PT-A & B in 6 hours Reliability = .90	MBE weighted 50% 3 Essays in 3 hours 1 PT in 3 hours Reliability = .84	93%
Models 2a and 2b cut test length in half with MBE weighted 50%		

Model 2a	Model 2b	% Agree
MBE weighted 50% Essays 1-3 in 3 hours PT-A in 3 hours Reliability = .85	MBE weighted 50% Essays 4-6 in 3 hours PT-B in 3 hours Reliability = .83	85%
Same models but completely different written questions in 6 hrs		

Model 3a	Model 3b	% Agree
MBE weighted 50% Essays 1-4 in 4 hours PT-A in 3 hours Reliability = .87	MBE weighted 50% Essays 3-6 in 4 hours PT-B in 3 hours Reliability = .86	88%
Models share 2 of their 4 essay questions in 7 hrs		

Model 4a	Model 4b	% Agree
MBE weighted 50% Essays 1-5 in 5 hours PT-A in 3 hours Reliability = .88	MBE weighted 50% Essays 2-6 in 5 hours PT-B in 3 hours Reliability = .88	91%
Proxy for a 6½ hour written exam (4 essay questions in common)		

Table 6
Side-By-Side Model Comparison Chart

Total testing time	3 days			2 Days
Written components	6 Essays + 2 PTs			3-4 Essays + 1 PT
Model	Model 1a	Model 1b		Models 2 & 3
Written/MBE weight	65/35	50/50		50/50
Score Reliability				
All Takers	.83	.88		.82 - .84
February	.81	.87		.80 - .82
July	.85	.90		.83 - .87
February Passing Rates				
All February takers	37%	37%		37%
Females	39%	37%		37%
Males	35%	37%		37%
White	41%	42%		42%
Asian	35%	35%		35%
Hispanic	28%	28%		28%
African American	20%	20%		21%
July Passing Rates				
All July takers	53%	54%		54%
Females	55%	54%		54%
Males	52%	54%		54%
White	60%	61%		61%
Asian	49%	49%		49%
Hispanic	40%	40%		41%
African American	24%	25%		25%

Total testing time includes the MBE. Models 2a and 2b use three 1-hour essay questions. Models 3a and 3b use four 1-hour essay questions. Results are based on a single reading of answers on the 10 February and 10 July exams given between 2001 and 2010 (total N = 125,178 candidates). Model 4's February and July passing rates were consistent with Model 1's rates.

ACS VENTURES
BRIDGING THEORY & PRACTICE

Conducting a Content Validation Study for the California Bar Exam

Final Report

October 4, 2017

Submitted By:

Chad W. Buckendahl, Ph.D.
702.586.7386
cbuckendahl@acsventures.com

Contents

Executive Summary....3
Introduction....5
Current Examination Design....5
Study Purpose....5
Procedures....6
Panelists....6
Workshop Activities....7
Orientation....7
Content Validity Judgments....8
Analysis and Results....9
Content Sampling Across Years....12
Evaluating the Content Validation Study....14
Procedural....15
Internal....16
External....16
Utility....17
Process Evaluation Results....18
Gap Analysis....19
Conclusions and Next Steps....21
References....23
Appendix A – Panelist Information....24
Appendix B – Content Validation Materials and Data....25
Appendix C – Evaluation Comments....26

Executive Summary

The California Bar Exam recently undertook a content validation study to evaluate the alignment of content and cognitive complexity on their exams to the results of a national job analysis. This study involved gathering judgments from subject matter experts (SMEs) following a standardized process for evaluating examination content, discussing judgments made by the SMEs, summarizing these judgments, and evaluating the representation of content on the examination.

In this process, content validation judgments for the assessments were collected on two dimensions – content match and cognitive complexity. The Written and Multistate Bar Exam (MBE) components of the examination were evaluated for their match to the results of the National Conference of Bar Examiners' (NCBE) 2012 job analysis in terms of content and cognitive complexity as defined by an adaptation of Webb's (1997) Depth of Knowledge (DOK). For the constructed response items (i.e., essay questions, performance task), score points specified in the scoring rubric were evaluated separately to acknowledge the potential for differential alignment evidence (i.e., that different aspects of the scoring criteria may measure different knowledge, skills, or abilities). Because MBE items were not available for the study, the subject areas as described in the publicly available content outline were reviewed and evaluated based on their proportional contribution to the examination.

Summary results suggested that all content on the examination matched with job-related expectations for the practice of law. The cognitive complexity for the written component of the examination as measured by DOK was also consistent with the level of cognitive complexity (e.g., analysis vs. recall) expected of entry-level attorneys. In addition, a review of the content sampling of the examination over time suggests that most content on the examination is consistent with content expected for entry level practice. The sampling plan and the current representation of knowledge and skills when considering the combination of the Written and MBE components of the examination suggest stable representation year to year. This is discussed in more detail in the body of the report. However, there are opportunities for improvement in both the content representation and sampling plan of the existing subject areas.

Results from the judgment tasks and qualitative feedback from panelists also suggested some formative opportunities for improvement in the structure and representation of content on the examination that could be considered. As recommended next steps for the California Bar Examination in its evaluation of its design and content, the results of the gap analysis and feedback from panelists provide a useful starting point for further discussion. Specifically, from the results of the national survey, skills and tasks were generally interpreted as more generalizable than many of the knowledge domains. Given the diversity of subject areas in the law, this is not surprising. At the same time, it may also suggest that a greater emphasis on skills could be supported in the future. To answer this question, further study is warranted. This additional study would begin with a program design that leads to a job analysis for the practice of law in California. As an examination intended to inform a licensure decision, the focus of the measurement of the examination needs to be on practice and not on the education or training programs. Through this combination of program design and job analysis, results would inform and provide evidence for decisions about the breadth and depth of measurement on the examination along with the relative emphasis (e.g., weighting) of different components.

While the results of this study provided evidence to support the current iteration of the examination, there are also formative opportunities for the program to consider in a program redesign. Specifically, the current design and format for the California Bar Examination has been in place for many years. Feedback from the content validation panelists suggested that there are likely subject areas that could be eliminated or consolidated to better represent important areas needed by all entry-level practitioners. From a design perspective, it may be desired to define the components of the examination as a combination of a candidate's competency in federal law, California-specific law, and job-related lawyer skills. Further, if the MBE continues to be included as part of the California Bar Examination, it would be important to be able to review the items on a recently operational form (or forms) of the test to independently evaluate the content and cognitive complexity of the items. If the California is unable to critically review this component of their program, it should prompt questions about whether it is appropriate to continue to include it as part of their examination.

Similarly, such a redesign activity would offer the program an opportunity to evaluate the assessment item types of the examination (e.g., multiple choice, short answer, extended response), scoring policies and practices for human scored elements (e.g., rubric development, calibration, evaluation of graders), alternative administration methods for components (e.g., linear on the fly, staged adaptive, item level adaptive), and alternative scoring methods for constructed response (e.g., automated essay scoring). Advances in testing practices and technologies as well as the evolution of the practice of law since the last program design activity suggest that this interim study may facilitate additional research questions. As an additional resource about the current practices within credentialing programs, interested readers are encouraged to consult Davis-Becker and Buckendahl (2017) or Impara (1995).

For licensure examination programs, in terms of evidence to define content specifications, the primary basis for evidence of content validity come from the results of a job analysis that provides information about the knowledge, skills, and abilities for entry-level practitioners. Although the results of the 2012 NCBE job analysis were used for this study, it would be appropriate for the program to conduct a state-specific study as is done for other occupations in California to then be used to develop and support a blueprint for the examination. The specifications contained in the blueprint are intended to ensure consistent representation of content and cognitive complexity across forms of the examination. This would strengthen the content evidence for the program and provide an opportunity for demonstrating a direct link between the examination and what occurs in practice. These two activities – program design and job analysis – should be considered as priorities with additional redevelopment and validation activities (e.g., content development, content review, pilot testing, psychometric analysis, equating) occurring as subsequent activities.

Recognizing the interrelated aspects of validation evidence for testing programs, it is valuable to interpret the results of this study and its potential impact on the recently conducted standard setting study for the California Bar Examination. Specifically, the results of the content validation study suggested that most of the content on the examination was important for entry level practice without substantive gaps in what is currently measured on the examination compared with what is expected for practice. However, if the examination is revised in the future, it would likely require revisiting the standard setting study.

The purpose of this report is to document who was involved in the process, processes that were used, results of the content validation study, conclusions about content validity of the examination, and recommendations for next steps in the examination development and validation process.

Introduction

The purpose of licensure examinations like the California Bar Exam is to distinguish candidates who are at least minimally competent from those that could do harm to the public (i.e., not competent). This examination purpose is distinguished from other types of exams in that licensure exams are not designed to evaluate training programs, evaluate mastery of content, predict success in professional practice, or ensure employability. As part of the validation process for credentialing examinations, a critical component includes content validation (see Kane, 2006). Content validation involves collecting and evaluating evidence of alignment of content (e.g., knowledge, skills, abilities) and cognitive processing (e.g., application, analysis, evaluation) to established job-related knowledge, skills, abilities, and judgments. Substantive overlap between what is measured by the examination and what is important for entry level practice is needed to support an argument that the content evidence contributes to valid scores and conclusions.

Current Examination Design

The California Bar Exam is built on multiple components intended to measure the breadth and depth of content needed by entry level attorneys. Beginning with the July 2017 examination, these components include the Multistate Bar Exam (MBE) (175 scored and 25 unscored multiple-choice questions), five essay questions, and a performance task. The combined score for the examination weights the MBE at 50% and the written response components at 50% with the performance task being weighted as twice as much as an essay question.[1] A decision about passing or failing is based on the compensatory performance of applicants on the examination and not any single component. This means that a total score is used to make decisions and no one question or task is determinant of the pass/fail determination.

Study Purpose

The purpose of this study was to evaluate the content representation and content complexity of the California Bar Examination in comparison with the results of a job analysis conducted by the National Conference of Bar Examiners (NCBE) in 2012. To collect the information to evaluate these questions, Dr. Chad Buckendahl of ACS Ventures, LLC (ACS) facilitated a content validation workshop on June 6-8, 2017 in San Francisco, CA. The purpose of the meeting was to ask subject matter experts (SMEs) to make judgments about the content and cognitive complexity of the components of the California Bar examination.

This report describes the sources of validity evidence that were collected, summarizes the results of the study, and evaluates the results using the framework for alignment studies suggested by Davis-Becker and Buckendahl (2013). The conclusions and recommendations for the examination program are based on this evaluation and are intended to provide summative (i.e., decision making) and formative (i.e., information for improvement) feedback for the California Bar Examination.

[1] Before July 2017, the written section of the bar exam was weighted 65 percent of the total score and consisted of six essay questions and two performance test questions administered over two days.

Procedures

The content validation approach used for the study relies on the content and cognitive complexity judgments suggested by Webb (1997). In this method, panelists make judgments about the cognitive complexity and content fit of exam items or score points relative to content expectations. For this study, those content expectations were based on the 2012 NCBE job analysis supplemented by links to the U.S. Department of Labor's O*NET[2] regarding lawyers that was updated in 2017.

A job analysis is a study often conducted every five to seven years to evaluate the job-related knowledge, skills, and abilities that define a given profession. Conducting a job analysis study for a profession can often take 9-12 months to complete. In using the results from the NCBE study as a reference point, these data were within the typical range for conducting these studies and it was a readily available resource given the timeline under which the California Bar Exam was asked to provide evidence of content validation of its examination.

Panelists

Ten panelists participated in the workshop and were recruited to represent a range of stakeholder groups. These groups were defined as Recently Licensed Professionals (panelists with less than five years of experience), Experienced Professionals (panelists with ten or more years of experience), and Faculty/Educator (panelists employed at a college or university). A summary of the panelists' qualifications is shown in Table 1.

Table 1. Profile of content validation workshop panel

Race/Ethnicity	Freq.	Percent
Asian	1	10.0
Black	2	20.0
Hispanic	1	10.0
White	6	60.0
Total	10	100.0

Gender	Freq.	Percent
Female	5	50.0
Male	5	50.0
Total	10	100.0

Years of Practice	Freq.	Percent
5 Years or Less	2	20.0
>=10	8	80.0
Total	10	100.0

Nominating Entity	Freq.	Percent
ABA Law Schools	2	20.0
Assembly Judiciary Comm.	1	10.0
Board of Trustees	1	10.0
BOT – COAF[3]	3	30.0
CALS Law Schools	1	10.0
Registered Law Schools	1	10.0
Senior Grader	1	10.0

Employment type	Freq.	Percent
Academic	3	30.0
Large Firm	2	20.0
Non Profit	1	10.0
Small Firm	1	10.0
Solo Practice	3	30.0

[2] The O*NET is an online resource when evaluating job-related characteristics of professions. See https://www.onetonline.org/ for additional information.
[3] Council on Access & Fairness.

Total	10	100.0	Total	10	100.0

Workshop Activities

The California Bar Exam content validation workshop was conducted June 6-8, 2017 in San Francisco, CA. Prior to the meeting, participants were informed that they would be engaging in tasks to evaluate the content and cognitive complexity of the components of the California Bar Examination. The content validation process included an orientation and training followed by operational alignment judgment activities for each essay/performance task and MBE subject area, as well as written evaluations to gather panelists' opinions of the process. Workshop orientation and related materials are provided in Appendix B.

Orientation

The meeting commenced on June 6th with Dr. Buckendahl providing a general orientation and training for all panelists that included the goals of the meeting, an overview of the examination, cognitive complexity levels, and specific instructions for panel activities. Additionally, the orientation described how the results would be used by policymakers and examination developers to evaluate the current structure and content representation of the examination.

Specifically, the topics that were discussed in the orientation included:

- The interpretation and intended use of scores from the California Bar Exam (i.e., licensure)
- Background information on the development of the California Bar Exam
- Summary results of the NCBE job analysis and O*NET descriptions
- Purpose of alignment information for informing validity evidence

After this initial orientation, the panel was trained on the alignment processes that were used. This training included discussions of the following:

- Cognitive complexity framework – understanding each level, evaluating content framework
- Content match – evaluating fit of score points or subject areas to job-related content
- Decision making process – independent review followed by group consensus

After the training, the panelists began making judgments about the examination. Their first task involved making judgments about the intended cognitive complexity of the knowledge, skills, abilities, and task statements from the 2012 NCBE job analysis. The cognitive complexity framework used for this study was an adaptation of Webb's (1997) Depth of Knowledge (DOK) for a credentialing exam. The DOK levels represent the level of cognitive processing associated with performing a task or activity. Lower DOK levels correspond to cognitive processes such as recall or remembering while higher levels correspond to application of knowledge, analysis, or evaluation. Within Webb's (1997) framework, Level 1 is defined as recall and reproduction, Level 2 is defined as working with skills and concepts, Level 3 is defined as short-term strategic thinking, and Level 4 is defined as extended strategic thinking. For this study, Level 1 was defined as recall or memorization, Level 2 was further clarified as representing the understanding and application level of cognitive process, Level 3 was defined as analysis and evaluation, and Level 4 was defined as creation of new knowledge.

Within psychological measurement, the depth of cognitive processing is considered in combination with the content to ensure that the claims made about candidates' abilities are consistent with the target construct. The DOK framework is one of many potential scales that can be used to evaluate this aspect of content.

Another commonly used model comes from Bloom (1956) and defines cognitive processes being knowledge, comprehension, application, analysis, synthesis, and evaluation. The inclusion of cognitive complexity as a consideration in the evaluation of the content validity of the California Bar Exam is important because it provides information on not only *what* may be needed on the examination, but *at what cognitive level* should candidates be able to function with the content. Procedurally, after rating the DOK of the first few statements as a group, panelists made judgments independently followed by consensus discussions. This consensus judgment was then recorded and used for the subsequent analysis.

Content Validity Judgments

Although characterized as "content," content validation is inclusive of judgments about both cognitive complexity and content match. After a review of the knowledge and task statements from the job analysis, the panelists began reviewing the components of the examination. For these components, panelists made independent judgments regarding the content match with the results of the NCBE job analysis. To calibrate the group to the process and the rating tasks, some of the judgments occurred as a full group facilitated discussions with other judgments occurring independently followed by consensus discussions. At key phases of the process panelists completed a written evaluation of the process including how well they understood the alignment tasks, their confidence in their judgments, and the time allocated to make these judgments.

On the first day, panelists reviewed and determined the cognitive complexity levels of each knowledge and task statement of the job analysis. This activity was done to establish the expected depth of knowledge (DOK) associated with the respective knowledge, skill, and ability (KSA). A summary of the results from these judgments suggested that most KSAs were judged to be at Levels 2 and 3 of the DOK framework. This means that most of the California Bar Examination is expected to measure candidates' abilities at levels beyond recall and memorization, specifically at the understanding, application, analysis, and evaluation levels. As shown below in Table 2, the current examination illustrates measurement expectations consistent with these expectations.

On the second day, the panel began making alignment judgments on the essay questions with the first one occurring as a full group activity. This was then followed by dividing up the task to have two subgroups each evaluate two essay questions and come to consensus on the judgments. After completing judgments on the essay questions, the full group then reviewed the expected content and DOK for the performance task and discussed the representation of content/skills. The third day then involved a full group facilitated discussion where judgments about the representation of domains of the MBE examination to evaluate proportional contribution to the overall content representation.

These judgment activities were followed by a facilitated discussion about content that could be measured on the examination that was not discussed (e.g., subject areas that were measured in other years). A related part of this brief discussion was where content that is eligible for sampling on the California Bar Exam may be more appropriately represented (e.g., Bar Examination, MCLE). These results are included in the evaluation section of this report, but should not be interpreted as a program design or redesign activity. The inclusion of this part of the study responded to a request to gather some high-level information as a starting point for additional exploration of how the California Bar Examination should be defined and structured.

Analysis and Results

The content validation findings are intended to evaluate the following questions:

- What is the content representation of the California Bar Exam essay questions, performance task, and MBE subject areas relative to the knowledge and task statements of the NCBE job analysis?
- What knowledge and task statements from the NCBE job analysis are NOT covered by the California Bar Exam?
- What California Bar Exam content does NOT align with the knowledge and task statements of the NCBE's 2012 job analysis?

There are currently 13 subject areas that can be sampled on the written portion of the California Bar Exam. This means that not all subject areas can be included each year and need to be sampled over time. To answer these content validation questions, the proportional contribution (i.e., percentage) of each exam component was estimated to approximate the distribution of content for the examination. This distribution is influenced by the sampling of content that occurs on the examination each year. As noted, each of the 13 subject areas cannot be included each year, so the content specifications require sampling to occur over multiple years.

For example, if a Real Property essay question is included for an examination, we would expect to see greater representation of the Real Property subdomain relative to years where this subject area is not included as part of the sampling plan. This is also why consideration was given to the content sampling plan for the program and not any single year. To apply a content sampling approach, it is important that the examination meet an assumption of unidimentionality (i.e., there is a dominant construct that is measured by the exam). If this assumption is met, then the variability of content year-to-year does not pose a significant threat to the validity of interpretations of the scores, even if there is an intuitive belief about what content should or should not be on the examination.

To illustrate the effect of the content sampling over time, it is important to understand what parts of the examination are constant versus variable across years. With the weighting of the exam beginning in July 2017 being 50% from the Multistate Bar Exam (MBE) and 50% from the written component (i.e., essay questions and performance task, we can calculate how much each part of the examination contributes to the whole. This breakdown is shown here:

Multistate Bar Exam (50%)

- The MBE is comprised of seven subject area sections, each with 25 scored questions. This means that each of these sections contributes approximately 7% to the total score (i.e., 1 section divided by 7 total sections and then multiplied by 50% to reflect that the MBE is only half of the exam).
- The blueprint for the MBE is fixed, meaning that the same seven content areas are measured each year. Therefore, the representation of content from this exam is consistent year-to-year until any changes are made to the blueprint.

Written Component (50%)

- The written component of the examination is comprised of five essay questions and one performance task that is weighted twice as much as one essay question. This means that for the written component, each of the five essay questions represent approximately 7% of the total score and the performance task represents approximately 14% of the total score (i.e., 1 essay

question divided by 7 total scoring elements [the performance task is calculated as 2 divided by 7 total scoring elements to reflect the double weighting] and then multiplied by 50% to reflect that the written component is only half of the exam.

- The blueprint for the written component is fixed with respect to the number of essay questions and performance task, but there is content sampling that occurs across the 13 subject areas currently eligible for selection. However, one of these subject areas, Professional Responsibility, is represented each year on the examination. Additional discussion about the potential impact of content sampling is discussed below.

The summary matrix in Table 2 represents the combination of information from the cognitive complexity ratings (reflected as Depth of Knowledge levels) in addition to the proportion of aligned content. For efficiency, the results are included for areas of content that were judged to align. Note that there were no components or subcomponents of the California Bar Exam that did not align with knowledge and task statements from the job analysis. There were, however, some areas suggested by the job analysis that could be considered in future development efforts by the Bar Exam that are discussed in the Conclusions and Next Steps section of this report.

Additional explanation is needed for readers to interpret the information presented in Table 2. Within the table, the first two columns refer to the knowledge, skills, abilities, or general tasks that were part of the summary results from the NCBE job analysis. Information in the third column relies on a coding scheme where K-1 refers to the first knowledge statement, S-1 refers to the first skill statement, A-1 refers to the first ability statement, T-1 refers to the first task statement in the O*NET framework. Other links within this framework will associate a letter and numerical code to the appropriate statement (e.g., K-2 refers to the second knowledge statement, T-3 refers to the third task statement). This information is provided to illustrate alignment with a concurrent source of evidence regarding knowledge, skills, abilities, and tasks that may be representative of entry-level practice. For interested readers, the narrative descriptions of these links to the O*NET that were used by panelists are provided in Appendix B.

The Statement DOK column provides information about the expected cognitive complexity for entry-level lawyers on the given knowledge, skill, ability, or task statement with lower numbers being associated with lower levels of cognitive complexity on the 1 (recall or memorization), 2 (understanding and application), 3 (analysis and evaluation) and 4 (creation) scale described above.

In the last three columns of Table 2, information about the estimated percent of the examination that was represented by content on the July 2016 administration with an important caveat. Because the goal of the content validation study was to evaluate the content representation that may occur on the California Bar Examination based on the new examination format that began in July 2017, we selected five essay questions and a performance task as representative of how an examination could be constructed without regard to specific content constraints (i.e., specific subject areas that may be included). This means that the interpretation of the results is dependent on the content sampling selected for the study. This concept is further discussed in the next section.

As described above, to calculate the percentage of coverage for a given content area, we first applied the weights to the respective components of the examination (i.e., 50% for the essays and performance task [written] component, 50% for the multistate bar exam [MBE]). We then calculated the proportion of each subsection within a component based on its contribution to the total score. For example, each essay question

is weighted equally with the performance task weighted twice as much as an essay question. This means that within the written component, there are six questions where one of the questions is worth twice as much. Proportionally, this means that each essay question is worth approximately 14% of the written component score whereas the performance task is worth approximately 28% of the written component score.

However, because the written component only represents half of the total test score, this means that these percentages are multiplied by 50% to determine the weight for the full examination (i.e., approximately 7% for each essay question, 14% for the performance task). The same calculation was applied to the seven equally weighted sections of the MBE. Ratings from panelists on each of the essay questions, performance task, and the content outline from the MBE were communicated as consensus ratings and based on proportional contributions of knowledge, skills, and abilities. These proportions could then be analyzed as weights based on the calculations described above to determine the component and overall content representation.

Table 2. Consolidated content validation results with approximate percentage of representation.

				% of Exam		
	Knowledge, Skills, Abilities and Tasks from the NCBE Job Analysis Survey	**Link to O*NET**	**Statement DOK**	**Essays and PT**	**MBE**	**Total**
Section I. Knowledge Domains[4]						
1	Rules of Civil Procedure	K-1	2	4%	4%	7%
2	Other Statutory and Court Rules of Procedure	K-1	1	-	4%	4%
3	Rules of Evidence	K-1	2	-	7%	7%
10	Contract Law[5]	K-1	2	3%	7%	10%
11	Tort Law	K-1	2	-	4%	4%
12	Criminal Law	K-1	2	-	7%	7%
13	Rules of Criminal Procedure[6]	K-1	2	-	-	0%
14	Other Privileges[7]	K-1	2	-	-	0%
15	Personal Injury Law	K-1	1	-	4%	4%
19	Principles of Electronic Discovery[8]	K-1	1	1%	-	1%
20	Real Property Law	K-1	2	3%	7%	10%
21	Constitutional Law[9]	K-1	2	3%	7%	10%

[4] Note that a current content constraint of the examination is that Professional Responsibility and Ethics is represented on each form of the test. When this content area is included it would reduce the representation of another content area that would be sampled.
[5] MBE content for this area was also judged to partially align with Real Property.
[6] MBE content for this area was also judged to partially align with Criminal Law and Procedure.
[7] MBE content for this area was also judged to partially align with Evidence.
[8] MBE content for this area was also judged to partially align with Civil Procedure.
[9] MBE content for this area was also judged to partially align with Civil Procedure, Criminal Law and Procedure, and Torts.

24	Family Law	K-1	2	3%	-	3%
Section II. Skills and Abilities						
87	Written communication	S-9, A-5	3	4%	-	4%
93	Critical reading and comprehension	S-3, A-3	3	3%	-	3%
94	Synthesizing facts and law	A-7	3	8%	-	8%
95	Legal reasoning	A-6, A-7	3	15%	-	15%
100	Issue spotting	S-5	3	1%	-	1%
108	Fact gathering and evaluation	S-5	3	2%	-	2%
Section III. General Tasks						
123	Identify issues in case	T-1, T-12	2	2%		2%
	Total			50%	50%	100%[10]

As shown in the footnotes of Table 2, there were areas of the MBE that could represent additional areas of content. However, the extent of that alignment is unknown because we did not have access to the actual test items; only the publicly available content outline. As a result, this report includes the judgments from the panel as a reference point for future study if the actual forms of the MBE are available for external evaluation in the future. To avoid speculation for this report, we did not estimate the potential contribution of these additional areas and only noted them.

Content Sampling Across Years

As noted above, the written component of the examination currently samples from 13 subject areas. Table 3 shows the number of times that each of these subject areas has been represented by essay questions over the last decade. This information is useful to evaluate whether the content emphasis is consistent with the subject areas that have been judged as more or less important in the practice analysis. In noting that one of the subject areas, Professional Responsibility, is sampled every year, we would expect some variability in the other four essay questions as subjects are sampled across years. Note that the performance is not related to the subject area and focuses specifically on lawyer skills, so the proportional measurement of these abilities also appears to be consistent across years.

[10] Note that totals for each component of the examination and overall will not equal 100% due to rounding.

Table 3. Representation of subject areas from 2008-2017 (n=20 administrations).

Subject area	Frequency of representation[11]	Rating of significance[12]	Percent Performing[13]
Professional Responsibility	19	2.83	93%
Remedies	12	N/A[14]	N/A
Business Associations	11	2.33	67%
Civil Procedure	10	3.08	86%
Community Property[15]	10	2.23	53%
Constitutional Law	10	2.29	76%
Contracts	10	2.67	84%
Evidence	10	3.01	81%
Torts	10	2.50	61%
Criminal Law and Procedures	9	2.50/2.47[16]	54%/54%
Real Property	9	2.30	56%
Trusts	7	1.95	44%
Wills	7	2.21	46%

[11] Frequency is defined as the number of times a subject area was represented as a main or crossover topic on the California Bar Examination from 2008-2017.
[12] Ratings are based on the average Knowledge Domain ratings for the 2012 NCBE Job Analysis study on a scale of 1 to 4 with values closer to 4 representing more significant content.
[13] Ratings are based on the percentage of respondents indicating that they perform the knowledge for the 2012 NCBE Job Analysis study. Values range from 0% to 100% with higher percentages indicating that more practitioners perform the knowledge.
[14] Remedies does not align with a single Knowledge Domain because it crosses over multiple, substantive areas of practice in law.
[15] Community Property was interpreted to be part of Family Law.
[16] Criminal Law and Procedures were asked as separate Knowledge Domain statements. Each significance rating is included.

For the essay questions in this study, panelists judged each one as measuring approximately 50% of the subject area knowledge (e.g., real property, contracts) and 50% of lawyer skills (e.g., application of law to facts, analysis, reasoning). This means that for a given essay question, the measurement of the subject area knowledge represents approximately 3.6% of the total examination (i.e., each essay question contributes approximately 7% to the total score (7.14% to be more specific), so if 50% of this is based on the subject area, 7% multiplied by 50% results in approximately 3.6% of the measurement being attributable to the subject area).

Knowing that the current sampling plan includes Professional Responsibility effectively yearly along with the performance task, this means that subject area sampling only applies to the four essay questions that may represent a different subject area year-to-year. In aggregate, this means that the potential variability in the measurement of the examination across years is approximately 14%-15% (i.e., 3.6% multiplied by the 4 essay questions). Another way to communicate these results is to say that 85%-86% of the measurement of the examination remains constant across years. This suggests that what is being measured on the examination remains stable.

In addition, the relationship between the emphasis of the subject areas in Table 3 as represented by the frequency of occurrence, the average significance rating, and the percent performing provides some information that will inform future examination redevelopment. Specifically, the correlation between the frequency of subject areas being represented on the examination and the average significance rating was 0.48 while the correlation between the frequency of subject area representation and percent performing was 0.70. The correlation between the significance of the topic and the percent performing was 0.83. However, these results should not be over-interpreted based on the limited number of observations (n=12). These results suggest that there is moderate relationship between the content sampling and evidence of importance of subject areas to entry level practice. However, there are likely opportunities to further align the content sampling with subject areas that were rated as more or less significant for entry-level practice.

Evaluating the Content Validation Study

To evaluate the content validation study, we applied Davis-Becker and Buckendahl's (2013) framework for alignment studies. Within this framework, the authors suggested four sources of evidence that should be considered in the validation process: procedural, internal, external, and utility. If threats to validity are observed in these areas, it will inform policymakers' judgments regarding the usefulness of the results and the validity of the interpretation. Evidence within each area that was observed in this study is discussed below.

One important limitation of the study that could pose a threat to the validity of the results is the lack of direct evidence from the MBE. Content validation studies generally involve direct judgments about the characteristics of the examination content. Because examination items (i.e., questions) from the MBE were not available for the study, panelists were asked to make judgments about the content evidence from publicly available subject matter outlines provided by the NCBE. There is then assumption that items coded to these sections of the outline align as intended. However, these assumptions should be directly reviewed. Because

California is using scores from the MBE as an increasingly important component of its decision-making process, it is reasonable to expect that NCBE make forms of the test available for validation studies.[17]

Procedural

Procedural evidence was available when considering panelist selection and qualifications, choice of methodology, application of the methodology, and panelists' perspectives about the implementation of the methodology. For this study, the panel that was recruited represented a range of stakeholders: both newer and more experienced attorneys as well as representatives from higher education. Because content validation judgments are more objective in nature (i.e., what does this question measure) as opposed to making standard setting judgments (e.g., how would a minimally competent candidate perform), there are fewer criteria needed with respect to panelist selection other than that they were knowledgeable about the content and familiar with the population of examinees. Again, this was not an activity to determine what *should* be on the examination, but rather, what *is* currently being measured by the examination.

In selecting the methodology for the study, alternative designs were considered. One design could have had panelists making judgments about whether the content and cognitive complexity of the components of the examination were appropriate for entry-level practice. The risk in this approach is the diverse opinions represented by stakeholder groups without a common reference point or link to evidence of what occurs in practice. This type of evidence is typically available following a practice analysis and is then used to build a blueprint from which examination forms are constructed. At that point, such a design could have been implemented because the common reference point would have been the blueprint that was developed with a clear link to practice. However, this information was not available; therefore, this design would have been inappropriate and would have only highlighted individual panelists' opinions or biases (e.g., practitioners' preference for content that aligns with their respective area(s) of practice, high education representatives' preference for content that aligns with their curriculum).

To have a common reference point for panelists to evaluate the alignment of content, we selected the summary results from the 2012 NCBE job analysis study. These results were derived from a national survey that collected information about the knowledge, skills, abilities, and tasks of lawyers. Although the results were not specific to California, it is reasonable to expect that these results would generalize to expectations for attorneys in California. So, the design that included this information along with the evidence from the U.S. Department of Labor's O*NET provided concurrent evidence of the characteristics of attorneys in practice.

For the rating activities, essay questions and the performance task are based on scoring considerations that include multiple traits. Therefore, panelists were asked to breakdown the scoring to proportionally align the parts of these questions that matched with different knowledge, skills, or abilities. To have only evaluated the questions holistically would not have revealed the differential content representation. Given the constructed response aspects of the essay questions and performance task, the methodology and rating tasks were consistent with the types of questions and judgments that could be provided.

[17] For security reasons and to protect the integrity of the empirical characteristics of operational questions, NCBE only makes available practice questions or "retired" questions, but not the entire exam from a specific administration.

With respect to the process evaluation, panelists' perspectives on the process were collected and the evaluation responses were consistently positive suggesting that they understood the process and were confident in their judgments about the content validity. In addition, panelists provided comments about aspects of the process that could be improved. This feedback did not threaten the validity of the results, but does inform some of the suggested next steps for the program.

Internal

The internal evidence for content validation studies can be evaluated by examining the consistency of panelists' ratings and the convergence of the recommendations. One approach to content validity studies is to use one or more rating scales where panelists rate individual questions or score points on different criteria (Davis-Becker & Buckendahl, 2013). Decision rules can then be applied to analyze and evaluate the results along with calculating levels of agreement among the panelists. However, this methodology is often more appropriate with more discrete items.

For this study, the rating tasks and decision rules were based on consensus judgments that occurred based on discussions among panelists following individual ratings. This approach is more qualitative in nature and was selected based on the types of assessment items and corresponding scoring criteria/rubrics that were evaluated (i.e., constructed response) along with the lack of an opportunity for direct judgments on items on the MBE. Although the results should not be interpreted as unanimous support by the panelists, consensus was achieved for the content and cognitive complexity rating tasks.

External

The primary source of external evidence for the study was based on the results of 2012 NCBE job analysis as an indicator of suggested content for entry level practice based on a nationally representative sample of practitioners. In addition, links to the U.S. Department of Labor's O*NET that was updated for lawyers in 2017 were also included to provide another source. The summary results of the NCBE job analysis study included ratings of knowledge, skills, abilities, and tasks.

There is an important caveat to note about NCBE's study. Specifically, because the study was designed and implemented as a task inventory (i.e., a list of knowledge, skills, abilities, and tasks) rather than competency statements, there were many statements that were redundant, overlapping, or that could be consolidated or subsumed within other statements. This means that an activity such as preparing a memo for a client was broken down into its component parts (e.g., critical reading and comprehension, identifying the primary question, distinguishing relevant from irrelevant facts, preparing a written response) were listed as separate statements when these part of the same integrated, job related task. More important, the scoring criteria or rubric would not distinguish these elements and would instead allocate points for skills such as identifying and applying the appropriate legal principles to a given fact pattern or scenario; or drawing conclusions that are supported with reasoning and evidence.

However, the value of the job analysis study is that it served as a common, external source against which to evaluate the content and cognitive complexity of the California Bar Examination. A lack of overlap in some areas should not be interpreted as a fatal flaw due to the design of the job analysis. The results can be used to inform next steps in evaluating validity evidence for the program.

Utility

Evidence of utility is based largely on the extent to which the summative and formative feedback can be used to inform policy and operational decisions related to examination development and validation. The summative information from the study suggests that the content and cognitive complexity as represented by content of the examination are consistent with expectations for entry level attorneys when compared with the highly rated knowledge, skills, and abilities of the 2012 NCBE job analysis.

However, whether the proportional contribution of this content (i.e., the percentage of representation of the range of knowledge, skills, abilities) is being implemented as intended is a question that would need to be evaluated as part of the next steps for the program. The intended representation of content for a credentialing examination is generally informed by a job analysis (also sometimes called a practice analysis or occupational analysis, see Clauser and Raymond (2017) for additional information).

These studies often begin with a focus group or task force that defines the knowledge, skills, and abilities for the target candidate (e.g., minimally competent candidate, minimally qualified candidate) to create task or competency statements. These statements are then typically compiled into a questionnaire that is administered as a survey of practitioners to evaluate the relative emphasis of each statement for entry level practice. The results from the survey can then be brought back to the focus group or task force to discuss and make recommendations to the appropriate policy body about the recommended weighting of content on the examination. This weighting is communicated through an examination blueprint that serves as the guide for developing examinations for the program.

The formative information from the panelists' ratings for the individual essay questions and performance task can be evaluated internally to determine whether this is consistent with expectations. For example, if the panelists judged a question to require a candidate to demonstrate knowledge of a subject area as representing 50% of the measurement the question with the other 50% representing skills, the internal evaluation would ask the question of whether this was intended. This intent is evaluated through the design of the question, the stimulus material contained in it, the specific call of the question for the candidate, and the scoring criteria or rubric associated with the question. The information from this study provided evidence to the program of what is currently being measured by the California Bar Examination, but does not conclude whether this is the information that should be measured on the examination. That type of determination would be a combination of information from a job analysis in concert with discussions about the design.

In addition, the panelists' qualitative discussions about potential structural changes to the examination or whether some content is more appropriate as part of continuing education will be useful for policymaker deliberations and examination development purposes. The summary of this discussion is included as part of comments in Appendix C. However, because this was not a primary goal of the study, this information should be interpreted as a starting point for further study and evaluation, not for decision-making at this point. A program design activity that involves a look at the examination and the related components would be valuable to inform decision-making. For example, a potential design for the California Bar Examination might include the MBE as a measure of federal or cross-jurisdictional competencies, the essay questions may be useful for measuring subject areas of law that are important and unique to California, and the performance task serving as a content-neutral measure of the important skills that lawyers need in practice. However, this is a facilitated activity that is more appropriate for policymakers and practitioners to engage in as a precursor to the job analysis.

Process Evaluation Results

Panelists completed a series of evaluations during the study that included both Likert scale (i.e., attitude rating scale) and open-ended questions. The responses to the Likert scale questions are included in Table 4 and the comments provided are included in Appendix C. With respect to training and preparation, the panelists felt the training session provided them with an understanding of the process and their task. Following the training, the panelists indicated they had sufficient time to complete the rating process and felt confident in the results. The rating scales for questions can be interpreted as lower values being associated with less satisfaction or confidence with higher values being associated with greater satisfaction or confidence with the respective statement. Note that for question 2, panelists were only asked to indicate whether the time allocated for training was too little (1), about right (2), or too much (3).

Table 4. Summary of Process Evaluation Results

	Median	1	2	3	4
1. Success of Training					
Orientation to the workshop	4	0	1	3	6
Overview of alignment	4	0	1	3	6
Discussion of DOK levels	3.5	0	1	4	5
Rating process	3.5	0	1	4	5
2. Time allocation to Training	2	0	9	1	
3. Confidence in Cognitive Complexity Ratings	3	0	1	7	2
4. Time allocated to Cognitive Complexity Ratings	3	0	1	5	4
6. Confidence in Day 1 ratings	4	1	0	2	6
7. Time allocated to Day 1 ratings	3	0	0	5	4
9. Confidence in Day 2 ratings	3	0	0	5	3
10. Time allocated to Day 2 ratings	3.5	0	0	4	4
12. Confidence in Day 3 ratings	3.5	0	0	4	4
13. Time allocated to Day 3 ratings	3.5	0	0	4	4
14. Overall success of the workshop	3.5	0	0	4	4
15. Overall organization of the workshop	4	0	0	3	5

Gap Analysis

The content validation study was designed to evaluate the extent to which content on the California Bar Examination aligned with expectations for entry level practice for lawyers. In addition, a gap analysis was conducted to also respond to the question about what content may be important for entry level practice, but is not currently measured on the examination. For this analysis, two criteria were evaluated.

Specifically, the ratings of significance and percent performing from the NCBE job analysis survey were analyzed. For the purposes of this analysis, if a knowledge, skill, ability, or task (KSAT) statement received a significance rating of 2.5 or higher on a 1-4 scale, it was included as a potential gap. Note that some KSAT statements were not included, because they were ambiguous or not appropriate for the purposes of licensure (e.g., Professionalism, Listening Skills, Diligence). Further, statements that were judged to be subsumed within other statements (e.g., Organizational Skills as an element of Written Communication) are not included to avoid redundancy. The results of this analysis are shown in Table 5.

Table 5. Summary of gap analysis of content not primarily measured on the California Bar Examination.

	Knowledge, Skills, Abilities and Tasks from the NCBE Job Analysis Survey	Link to O*NET	Statement DOK	Significance (Mean)	% Performing
Section I. Knowledge Domains					
5	Research Methodology	K-1	2	2.91	89%
8	Statutory Interpretation	K-1	1	2.83	86%
9	Document Review/Documentary Privileges	K-1	2	2.73	81%
Section II. Skills and Abilities					
92	Using office technologies (e.g., word processing and email)	K-6	1	3.56	99%
102	Answering questions succinctly	N/A	1	3.30	99%
104	Computer skills	K-6	1	3.28	99%
105	Electronic researching	T-8	2	3.26	98%
113	Negotiation	S-7	1	2.97	87%
114	Resource management	K-4, T-11	1	2.93	96%
115	Interviewing	T-14	1	2.92	91%
118	Attorney client privilege - document reviewing	T-9	3	2.84	86%
119	Trial skills	T-7	1	2.71	68%
120	Legal citation	T-9, T-15	2	2.67	95%
Section III. General Tasks					
	Management of attorney-client relationship and caseload				
124	Establish attorney-client relationship	T-18	2	2.86	76%
125	Establish and maintain calendaring system	T-18	1	2.86	78%

127	Establish and maintain client trust account	T-21	1	2.52	36%
128	Evaluate potential client engagement	T-12	1	2.51	67%
	Research and Investigation				
142	Conduct electronic legal research	T-8	2	3.42	96%
143	Research statutory authority	T-8	2	3.38	95%
144	Research regulations and rules	T-8	2	3.31	96%
145	Research judicial authority	T-8	2	3.19	89%
146	Conduct document review	T-8	2	3.10	86%
147	Interview client and client representatives	T-14	2	3.04	77%
148	Conduct fact investigation	T-14	2	2.91	83%
149	Interview witness	T-14	1	2.75	69%
150	Research secondary authorities	T-8	2	2.70	92%
151	Obtain medical records	T-14	1	2.58	61%
152	Conduct transaction due diligence activities	T-2	1	2.54	58%
153	Request public records	T-16	1	2.53	81%
	Analysis and resolution of client matters				
157	Analyze law	T-1	3	3.46	97%
158	Advise client	T-2	2	3.20	87%
159	Develop strategy for client matter	T-13	1	3.13	87%
160	Negotiate agreement	T-9, T-10	1	2.93	77%
161	Draft memo summarizing case law, statutes, and regulations, including legislative history	T-15	3	2.81	86%
163	Draft demand letter	T-9	1	2.60	65%
164	Draft legal opinion letter	T-15	2	2.54	76%
165	Draft case summary	T-15	2	2.53	80%

The information from the gap analysis can be used to evaluate the current content representation of the examination to determine whether a) existing elements of measurement should be retained, b) new elements of measurement should be added, and c) the extent to which the current design of the examination supports measurement of the important aspects of the domain. A caution in interpreting these results is that some of the knowledge, skills, abilities, and tasks are not easily measurable in a written examination and may require different types of measurement strategies, some of these being potentially technology enhanced. An additional caution is that the statements from the 2012 NCBE job analysis overlapped with each other and were not mutually exclusive with respect to the tasks that lawyers might perform. For future studies, I would suggest a competency or integrated task statement based approach that is more consistent with the tasks, responsibilities, and activities that lawyers engage with as opposed to discrete aspects of practice.

Conclusions and Next Steps

At a summative level, the results of the content validation study suggest that the current version of the California Bar Examination is measuring important knowledge, skills, and abilities consistent with expectations of entry level attorneys as suggested by results from the 2012 NCBE job analysis. Whether the observed representation and proportional weighting are in alignment with the expectations for California cannot be determined without further evaluation. However, it is important to note that all content on the current examination was judged to align with elements of the NCBE job analysis that were rated as reasonably significant and/or performed frequently in practice. This also included the subject areas that are sampled across years, but were not included in this study.

As recommended next steps for the California Bar Examination in its evaluation of its design and content, the results of the gap analysis and feedback from panelists provide a useful starting point for further discussion. Specifically, from the results of the national survey, skills and tasks were generally interpreted as more generalizable than many of the knowledge domains. Given the diversity of subject areas in the law, this is not surprising. At the same time, it may also suggest that a greater emphasis on skills could be supported in the future. To answer this question, further study is warranted. This additional study would begin with a program design that leads to a job analysis for the practice of law in California. As an examination intended to inform a licensure decision, the focus of the measurement of the examination needs to be on practice and not on the education or training programs. Through this combination of program design and job analysis, results would inform and provide evidence for decisions about the breadth and depth of measurement on the examination along with the relative emphasis (e.g., weighting) of different components.

While the results of this study provided evidence to support the current iteration of the examination, there are also formative opportunities for the program to consider in a program redesign. Specifically, the current design and format for the California Bar Examination has been in place for many years. Feedback from the content validation panelists suggested that there are likely subject areas that could be eliminated or consolidated to better represent important areas needed by all entry-level practitioners.

To briefly reiterate an example described above, from a design perspective, it may be desired to define the components of the examination as a combination of a candidate's competency in federal law, California-specific law, and job-related lawyer skills. Further, if the MBE continues to be included as part of the California Bar Examination, it would be important to be able to review the items on a recently operational form (or forms) of the test to independently evaluate the content and cognitive complexity of the items. If the

California is unable to critically review this component of their program, it should prompt questions about whether it is appropriate to continue to include it as part of their examination.

Similarly, such a redesign activity would offer the program an opportunity to evaluate the assessment item types of the examination (e.g., multiple choice, short answer, extended response), scoring policies and practices for human scored elements (e.g., rubric development, calibration, evaluation of graders), alternative administration methods for components (e.g., linear on the fly, staged adaptive, item level adaptive), and alternative scoring methods for constructed response (e.g., automated essay scoring). Advances in testing practices and technologies as well as the evolution of the practice of law since the last program design activity suggest that this interim study may facilitate additional research questions. As an additional resource about the current practices within credentialing programs, interested readers are encouraged to consult Davis-Becker and Buckendahl (2017) or Impara (1995).

For licensure examination programs, in terms of evidence to define content specifications, the primary basis for evidence of content validity come from the results of a job analysis that provides information about the knowledge, skills, and abilities for entry-level practitioners. Although the results of the 2012 NCBE job analysis were used for this study, it would be appropriate for the program to conduct a state-specific study as is done for other occupations in California to then be used to develop and support a blueprint for the examination. The specifications contained in the blueprint are intended to ensure consistent representation of content and cognitive complexity across forms of the examination. This would strengthen the content evidence for the program and provide an opportunity for demonstrating a direct link between the examination and what occurs in practice. These two activities – program design and job analysis – should be considered as priorities with additional redevelopment and validation activities (e.g., content development, content review, pilot testing, psychometric analysis, equating) occurring as subsequent activities.

Recognizing the interrelated aspects of validation evidence for testing programs, it is valuable to interpret the results of this study and its potential impact on the recently conducted standard setting study for the California Bar Examination. Specifically, the results of the content validation study suggested that most of the content on the examination was important for entry level practice without substantive gaps in what is currently measured on the examination compared with what is expected for practice. However, if the examination is revised in the future, it would likely require revisiting the standard setting study.

References

Bloom, B., Englehart, M. Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, Toronto: Longmans, Green.

Clauser, A. L. & Raymond, M. (2017). Specifying the content of credentialing examinations. In S. Davis-Becker and C. Buckendahl (Eds.), *Testing in the professions: Credentialing policies and practice* (pp. 64-84). New York, NY: Routledge

Davis-Becker, S. & Buckendahl, C. W. (Eds.), (2017). *Testing in the professions: Credentialing policies and practice*. New York, NY: Routledge.

Davis-Becker S. & Buckendahl, C. W. (2013). A proposed framework for evaluating alignment studies. *Educational measurement: Issues and practice, 32*(1), 23-33.

Impara, J.C. (Ed.), (1995). *Licensure testing: Purposes, procedures, and practices*. Lincoln, NE: Buros Institute of Mental Measurements.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, CT: American Council on Education and Praeger.

Webb, N. L. (1997). Criteria for alignment of expectations and assessments in mathematics and science education (Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6). Madison, WI: University of Wisconsin, Wisconsin Center for Education Research.

Appendix A – Panelist Information

Content Validity Panelists.xlsx

Last Name	First Name	City	Role	Years in Practice
Baldwin-Kennedy	Ronda			
Barbieri	Dean			
Cramer	Mark			
Dharap	Shounak			
Gramme	Bridget			
Jackson	Yolanda			
Layon	Richard			
Lozano	Catalina			
Maio	Dennis			
Shultz	Marjorie			

Appendix B – Content Validation Materials and Data

The documentation used in the standard setting are included below.

Overview of Content Validation 1

Cal Bar Content Validation Worksho

Cal Bar Content Validation Worksho

Cal Bar Content Validation Worksho

NCBE Job Analysis Summary 2013

O*NET Summary for Lawyers

ACS VENTURES
BRIDGING THEORY & PRACTICE

California Bar Exam

Content Validation Workshop

Agenda

Tuesday, June 6

7:30 – 8:00	Breakfast
8:00 – 8:30	Introductions and Purpose of the Study
8:30 – 10:00	Initial training Purpose and design of the California Bar Exam Content validation judgments (Job Analysis/O*NET)
10:00 – 10:15	Break
10:15 – 11:45	DOK Ratings for knowledge, skills, and abilities (independent)
11:45 – 12:45	Lunch
12:45 – 2:15	DOK Ratings for knowledge, skills, and abilities (group consensus)
2:15 – 2:30	Complete first evaluation form
2:30 – 2:45	Break
2:45 – 3:45	Begin content validation judgments for first essay question (facilitated) Review scoring rubric/criteria for the question Evaluate content and cognitive complexity match
3:45 – 4:00	Break
4:00 – 4:45	Continue content validation judgments for first essay question (facilitated)
4:45 – 5:00	Complete second evaluation form

Wednesday, June 7th

8:00 – 8:30	Breakfast
8:30 – 9:30	Begin content validation judgments for second/fourth essay question (independent within subgroup) Review scoring rubric/criteria for the question Evaluate content and cognitive complexity match
9:30 – 10:15	Discuss initial content validation judgments for second/fourth essay question (subgroup)
10:15 – 10:30	Break
10:30 – 11:30	Continue content validation judgments for third/fifth essay question (independent)
11:30 – 12:15	Discuss initial content validation judgments for third/fifth essay question (subgroup)
12:15 – 1:00	Lunch
1:00 – 2:15	Begin content validation judgments for performance task (independent) Review scoring rubric/criteria for the question Evaluate content and cognitive complexity match
2:15 – 2:30	Break
2:30 – 3:30	Discuss initial validation judgments for performance task (group)
3:30 – 3:45	Break
3:45 – 4:45	Begin judgments for MBE Subject Matter Outline – content focus (independent)
4:45 – 5:00	Complete third evaluation form

Thursday, June 8

8:00 – 8:30	Breakfast
8:30 – 9:30	Continue judgments for MBE Subject Matter Outline – content focus (independent)
9:30 – 9:45	Break
9:45 – 10:45	Discuss judgments for MBE Subject Matter Outline (group)
10:45 – 11:00	Break
11:00 – 11:45	Continue discussing judgments for MBE Subject Matter Outline
11:45 – 12:00	Complete fourth evaluation form

Evaluation – 1 Content Validation Workshop

The purpose of this evaluation is to get your feedback about the various components of the content validation workshop. Please do not put your name on this evaluation form. The information from this evaluation will be used to improve future projects. Thank you!

Training

The training consisted of several components: orientation to the workshop, overview of alignment, discussion of cognitive complexity levels, and training on the rating process.

1. Using the following scale, please rate the success of each training component:

		Rating of Training Success			
Training Components		Very Unsuccessful			Very Successful
a.	Orientation to the workshop	1	2	3	4
b.	Overview of alignment	1	2	3	4
c.	Discussion of DOK levels	1	2	3	4
d.	Rating process	1	2	3	4

2. How would you rate the amount of time allocated to training?
   a. Too little time was allocated to training.
   b. The right amount of time was allocated to training.
   c. Too much time was allocated to training.

Cognitive Complexity Ratings of Job Analysis/O*NET KSAs

3. How confident were you about the cognitive complexity ratings you made?
   a. Very Confident
   b. Somewhat Confident
   c. Not very Confident
   d. Not at all Confident

4. How did you feel about the time available to make your cognitive complexity ratings?
   a. More than enough time was available
   b. Sufficient time was available
   c. Barely enough time was available
   d. There was not enough time available

5. Please provide any comments about the training or cognitive complexity ratings that would help in planning future workshops.

Evaluation – 2 Content Validation Workshop

Day 1 Content Validity Judgments

6. How confident were you about your Day 1 judgments of content validity for the California Bar Exam?
   a. Very Confident
   b. Somewhat Confident
   c. Not Very Confident
   d. Not at all Confident

7. How did you feel about the time allocated for making these judgments?
   a. More than enough time was available
   b. Sufficient time was available
   c. Barely enough time was available
   d. There was not enough time available

8. Please provide any comments about the Day 1 content validity activities that would be helpful in planning future workshops.

Evaluation – 3 Content Validation Workshop

Day 2 Evaluation of Essay Questions and Performance Task

9. How confident were you about your Day 2 judgments of content validity for the California Bar Exam?
   a. Very Confident
   b. Somewhat Confident
   c. Not Very Confident
   d. Not at all Confident

10. How did you feel about the time allocated for making these judgments?
   a. More than enough time was available
   b. Sufficient time was available
   c. Barely enough time was available
   d. There was not enough time available

11. Please provide any comments about the Day 2 rating activities that would be helpful in planning future workshops.

Evaluation – 4 Content Validation Workshop

Day 3 Evaluation of Content Outline for the MBE

12. How confident were you about your Day 3 judgments of content validity for the California Bar Exam?
    a. Very Confident
    b. Somewhat Confident
    c. Not Very Confident
    d. Not at all Confident

13. How did you feel about the time allocated for making these judgments?
    a. More than enough time was available
    b. Sufficient time was available
    c. Barely enough time was available
    d. There was not enough time available

Overall evaluation of the content validation workshop

14. Overall, how would you rate the success of the content validation workshop?
    a. Very Successful
    b. Successful
    c. Unsuccessful
    d. Very Unsuccessful

15. How would you rate the organization of the content validation workshop?
    a. Very Organized
    b. Organized
    c. Unorganized
    d. Very Unorganized

16. Please provide any comments about the content validation activities that would be helpful in planning future workshops.

Thank you for your contributions to the Content Validation Workshop!

Overview of Content Validation

Content serves as a prioritized source of validity evidence for credentialing (e.g., licensure, certification) examinations (see *Standards for Educational and Psychological Testing*, [AERA, APA, & NCME, 2014]). The process of content validation involves collecting and evaluating evidence alignment of content and response processes (e.g., cognitive complexity, depth of knowledge) with job-related knowledge, skills, abilities, and tasks. Substantive overlap between what is measured by an examination and what occurs in entry level practice is needed to support an argument that the content evidence contributes to valid scores and decisions for the purpose of licensure.

In the content validation study being conducted June 6th-8th, panelists will provide a series of judgments about the evidence of content and response processes for the California Bar Examination. The materials will include results from the most recent National Conference of Bar Examiners (NCBE) job analysis, information from the U.S. Department of Labor's O*NET, exam questions and scoring criteria from the 2016 exam, and the content outline from the Multistate Bar Examination (MBE). The tasks will involve making judgments about:

- Cognitive complexity/Depth of Knowledge – level of response processes for job analysis statements and elements of examination content; and
- Content – fit of score points or subject area topics to job-related content.

Procedurally, these judgments will occur in two phases. Panelists will initially make these judgments independently followed by consensus discussions with the group. This consensus judgment will be recorded by a table lead and used for the analysis.

The findings from the study will be used to evaluate several questions of alignment:

- What is the representation of content and cognitive level of the California Bar Examination score points relative to the knowledge, skill, and task statements of the NCBE job analysis?
- What knowledge and task statements from the NCBE job analysis are NOT covered by the California Bar Exam?
- What California Bar Exam content does NOT align with the knowledge and task statements of the NCBE job analysis?

Following the study, we will prepare a technical report that includes a summary of the alignment findings and results, including evidence of the people, process, results, and decision rules applied during the study.