



The State Bar *of California*

OPEN SESSION AGENDA ITEM 705 MAY 2020

DATE: May 14, 2020

TO: Members, Board of Trustees

FROM: Dag MacLeod, Chief of Mission Advancement & Accountability Division
Amy Nuñez, Director, Office of Admissions

SUBJECT: Report on and Approval of Recommendations Regarding the California Bar Exam Studies

EXECUTIVE SUMMARY

In 2017 the Supreme Court directed the State Bar to undertake studies of the California Bar Exam (CBX) to include “identification and exploration of all issues affecting California bar pass rates.” During the course of these studies it became clear that additional research was needed to ensure the reliability, validity, and fairness of the Exam. The Board of Trustees directed State Bar staff to undertake that research in its January 2018 update to the State Bar’s 2017–2022 Strategic Plan. Related, in 2017 the Governance in the Public Interest Task Force issued its triennial report. That report included recommendations for study and potential modification of certain Bar Exam processes, work that was recently completed. Finally, the release of bar exam topics prior to the July 2019 administration of the bar exam alerted the State Bar to the need to augment the various studies underway with an audit of the State Bar’s test administration processes and procedures.

This agenda item summarizes the findings and recommendations of this expansive body of work. The four studies addressed here include: the report of the California Attorney Practice Analysis Working Group; the Differential Item Functioning Analysis Report; the Review of the California Bar Examination Administration and Related Components; and A Report on the Phased Grading of the California Bar Examination. The agenda item then articulates an organizational framework for understanding the relationship among the four studies, and makes proposals for addressing the recommendations contained within them.

BACKGROUND

In 2017 the State Bar completed two major studies related to the California Bar Examination (CBX): a standard setting study to assess whether the pass score was set appropriately to reflect the minimum competence of entry-level attorneys, and a content validation study to evaluate whether the topics covered on the bar exam aligned with the knowledge, skills, and abilities required of entry-level attorneys.¹

While these studies represented important milestones in comprehensively assessing two critical elements of the CBX, the scope of the studies and the conclusions they reached were constrained by several factors. Most significantly, the content validation study relied heavily on a national survey of practicing attorneys conducted by the National Conference of Bar Examiners (NCBE) in 2012.² The national focus and age of that study resulted in a lack of current, California-specific, content for the evaluation of the written component of the CBX.

To remedy this, in 2018 the Board of Trustees of the State Bar created the California Attorney Practice Analysis (CAPA) Working Group to work with specialists in the field of psychometrics and practice analysis to document the practice of law in California. This information would provide the California-specific data necessary to evaluate the link between bar exam content and current legal practice, and to create a blueprint – an outline of content coverage across legal topics and job responsibilities – for future selection of bar exam topics and question items.³

While the CAPA Working Group was collecting data and evaluating the practice of law in California, a number of other studies on the bar exam were initiated.

Report on Differential Item Functioning Analysis⁴

As part of its commitment to access and inclusion in the legal profession, the State Bar initiated an evaluation of the fairness of the bar exam by evaluating the potential differential impact of exam questions by race, gender, and other factors. The statistical technique applied to an assessment of individual questions on an exam is referred to as Differential Item Functioning (DIF) analysis.

DIF analysis evaluates whether different groups perform consistently better or worse on specific questions. DIF analysis relies on Multistate Bar Exam (MBE) test scores to control for the underlying abilities of test takers. Items or questions are flagged as showing signs of DIF when the analysis shows statistically significant performance differences for test takers across

¹ For a summary of these reports, see Report to the Supreme Court of the State of California Final Report on the 2017 California Bar Exam Studies, December 1, 2017, <http://www.calbar.ca.gov/Portals/0/documents/reports/2017-Final-Bar-Exam-Report.pdf>

² The National Conference of Bar Examiners produces the licensing tests used by most US jurisdictions for admission to the State Bar including the Multistate Bar Examination (MBE), the Multistate Essay Examination (MEE), and the Multistate Performance Test (MPT).

³ See Attachment A, “DRAFT - The Practice of Law in California: Findings from the California Attorney Practice Analysis and Implications for the California Bar Exam.”

⁴ See Attachment B, “State Bar of California: Differential Item Functioning Analysis Report.”

groups that are otherwise comparable. Findings of DIF help to identify potential issues of bias in a test item.

Review of California Bar Examination Administration⁵

Shortly after the incident in July 2019 related to the disclosure of bar exam topics, the State Bar requested that the California Department of Consumer Affairs (DCA), Division of Programs and Policy Review (DPPR) conduct a review of the administration of the CBX. The review compared State Bar procedures and processes to best practices in testing as defined by the Standards for Educational and Psychological Testing (Standards).⁶

The DCA gathered and reviewed information on various aspects of the bar exam administration with the goal of identifying gaps and making recommendations to align State Bar practices more closely with the Standards. DCA staff reviewed operational manuals and procedures and conducted interviews with key staff from the Office of Admissions over a period of several months in late 2019 and early 2020.

The DCA's report concluded that "Overall, the State Bar is meeting professional guidelines and technical standards in most areas."⁷ Nonetheless, the report also provided observations and recommendations across a broad swath of CBX components that could be evaluated for further improvement, including registration for the bar exam, scoring, test security, instructions to test takers, and more.

Report on Phased Grading of the California Bar Examination⁸

The State Bar's 2017 Governance in the Public Interest Task Force (GTF) report included an extensive review of the work of various volunteer subentities of the State Bar including the Committee of Bar Examiners (CBE). Consultants engaged by the State Bar to conduct the CBE-related portion of the 2017 GTF review recommended that the CBE conduct a "review and evaluation of its grading process."

One key issue that quickly emerged from this evaluation was the question of whether the CBX could be graded more efficiently without sacrificing validity and reliability in the grading process. Typically it has taken, respectively, 11 and 16 weeks to finalize the grading of the February and July exams. Dr. Roger Bolus, the State Bar's longstanding, contract psychometrician, was asked to conduct this evaluation of the grading process with an emphasis on identifying efficiencies and best practices.

Combined, these studies present numerous opportunities for the State Bar to improve the operation of one of its most essential functions: evaluating candidates for admission to the practice of law. Moving from analysis and recommendations to policy development and implementation, however, requires the assessment of the four studies in relation to one another and in relation to other strategic considerations. Although some of the

⁵ See Attachment C, "Review of the California Bar Examination Administration and Associated Components."

⁶ American Educational Research Association, 2014

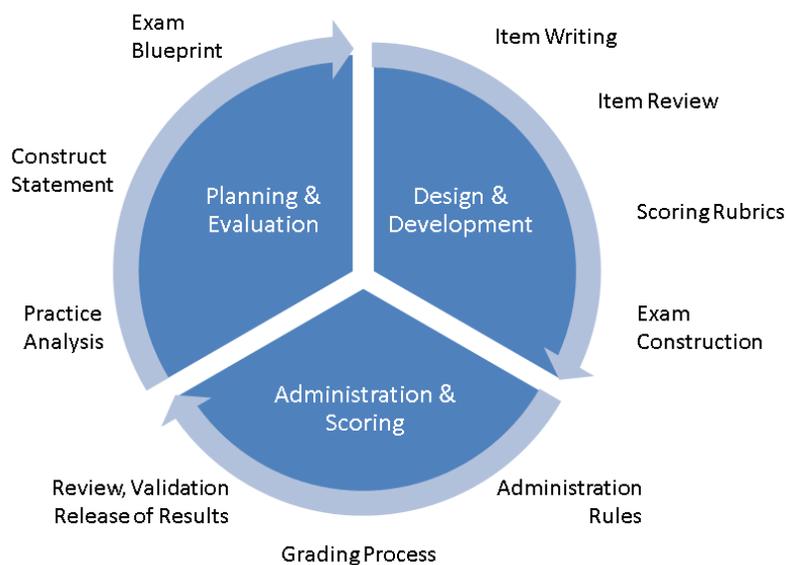
⁷ Page 30.

⁸ See Attachment D, "A Report on the Phased Grading of the California Bar Examination: A Profile of Recent Results and Modeling the Impact of Alternative Approaches."

recommendations might be implemented without affecting others, many of the recommendations – like many of the components of the CBX – are interdependent. And, while some of the recommendations are relatively limited in terms of their potential impact, others have far-reaching policy implications.

Figure 1 depicts the life cycle of a testing program, comprised of three distinct, interconnected phases: *Planning and Evaluation*, *Design and Development*, and *Administration and Scoring*. In addition to showing the interconnection among different components of the reports summarized here, Figure 1 also highlights the fact that a testing program needs to be continually updated so that a given exam reflects the actual practice of a profession.

Figure 1 – The Life Cycle of a Testing Program



The following section provides an overview of the key findings and recommendations of each of the four studies of the California Bar Exam. It then illustrates where the recommendations of each study fit within the three phases of the life cycle of a testing program. The organizational framework is then used to develop a roadmap to guide the implementation and further evaluation of the recommendations.

DISCUSSION

The presentation of the findings and recommendations from the bar exam studies proceeds in the order depicted in the life cycle of a testing program beginning with *Planning and Evaluation*. The CAPA study provides the foundation necessary for creating an exam blueprint. The Working Group with CAPA study oversight responsibility generated recommendations regarding the topics to be covered on the CBX. The report on Differential Item Functioning relates to the *Design and Development* phase of the life cycle, with important recommendations to ensure the fairness of the questions developed for the exam.

The Review of the California Bar Examination Administration report by the DCA is wide-ranging and straddles both the *Design and Development* phase and the *Administration and Scoring* phase of the life cycle. Finally, the Report on Phased Grading relates directly to the *Administration and Scoring* phase of the life cycle.

The California Attorney Practice Analysis (CAPA) Report

To address the limitations inherent in the 2017 reports on the bar exam, the State Bar committed to conducting a California-specific practice analysis. Practice analysis refers to “the systematic collection of data describing the responsibilities required of a professional and the skills and knowledge needed to perform these responsibilities.”⁹ The purpose of practice analysis for a licensing entity is twofold. First, practice analysis allows for the assessment of alignment between the content of a licensing exam and the profession being licensed. Second, practice analysis allows for the evaluation of the knowledge, skills, and abilities (KSAs) of people *at the entry level* of a profession.

The CAPA Working Group, established by the Board of Trustees, with members appointed by the Supreme Court, held its kick-off meeting in December 2018 with the goal of conducting a practice analysis for the State Bar of California. The Working Group was supported by and worked closely with Scantron, a consulting firm with deep expertise in the field of testing.

The practice analysis involved surveying over 125,000 active, licensed attorneys in California to collect data on attorney practices along two principal dimensions:

- What attorneys do as reflected in daily tasks; and
- What knowledge attorneys use to perform those tasks.¹⁰

Using two different, complementary survey methods, over 16,000 respondents provided information on how long they have been practicing law, the legal topics that they work on, and the skills that they use, along with ratings of the criticality of that work and the depth of knowledge required to perform it. These data were synthesized and analyzed by Scantron and State Bar staff and presented to the Working Group.

The research and analysis conducted by the CAPA Working Group coincided, and was coordinated, with a national practice analysis being conducted by the NCBE. The fact that the NCBE was updating its practice analysis provided another set of data points for the CAPA Working Group to consider in its deliberations. Although not planned from the outset, the concurrent NCBE practice analysis had the potential for allowing the Working Group to consider the question of whether California’s practice aligns sufficiently with national practice to recommend transitioning to the Uniform Bar Exam (UBE).¹¹

⁹ See “Practice Analysis: Building the Foundation for Validity,” Joan E. Knapp and Lenora G. Knapp, in *Licensure Testing: Purposes, Procedures, and Practices*, James C. Impara, editor, 1995.

¹⁰ A subgroup of the CAPA Working Group was created to draft detailed task descriptions and to construct a taxonomy of legal topics. The final list, used for data collection, consisted of 19 legal topics with 114 subtopics, and 13 competencies made up of 110 discrete tasks.

¹¹ The Uniform Bar Examination is composed of the Multistate Essay Examination (MEE), two Multistate Performance Test (MPT) tasks, and the Multistate Bar Examination (MBE) (note 2, above). “It is uniformly

The Working Group evaluated the findings and applied their professional judgment to determine whether specific areas of the law or skills should be tested on the CBX. Members of the Working Group engaged in robust debate over the relative merits of including or excluding specific areas of the law on the bar exam. In some cases, the Working Group concluded that a topic was too important to exclude, even if it was not commonly used by attorneys – Constitutional Law, for example. In other cases, the Working Group concluded that ensuring entry-level competence in a topic – Professional Competence, for example – might be better addressed through mandatory continuing legal education for first-year attorneys than inclusion on the Exam.

The Working Group’s consensus was to recommend that the State Bar:

- adopt the construct statement utilized by the Working Group to define the general scope of the bar exam and establish that the operational definition of how long an attorney may be considered an *entry-level attorney* is *within the first three years of practice*. The construct statement is as follows:

The California Bar Examination assesses legal knowledge, competency areas, and professional skills required for the entry-level practice of law and the effective, ethical representation of clients.

- adopt the following eight legal topics for a new bar exam content outline:
 - Administrative Law and Procedure
 - Civil Procedure
 - Constitutional Law
 - Contracts
 - Criminal Law and Procedure
 - Evidence
 - Real Property
 - Torts
- focus on the following broad areas of skills on the bar exam:
 - Drafting and writing;
 - Research and investigation;
 - Issue-spotting and fact-gathering;
 - Counsel/advice;
 - Litigation; and
 - Communication and client relationship

administered, graded, and scored and results in a portable score that can be transferred to other UBE jurisdictions,” see <http://www.ncbex.org/exams/ube/>

Although the Working Group reviewed the findings from the survey component of NCBE's job analysis, ultimately the timing of that national effort did not align with that of the CAPA Working Group. The NCBE's final report, to include recommendations regarding the topics and skills to be tested by the Multistate Bar Exam and the UBE, will not be issued until later this year.

Differential Item Functioning Analysis Report

As part of its commitment to ensuring fairness in all aspects of its admissions, regulation, and discipline work,¹² the State Bar hired a consultant from Scantron in the fall of 2019 to conduct a Differential Item Functioning (DIF) analysis for essay and performance test items on the bar exam. A test item or question is flagged as showing signs of DIF when the analysis indicates significantly different performance patterns for test takers across different groups with comparable underlying abilities.

The analysis looked at all essay and performance test (PT) questions used on the bar exam from July 2009 to February 2019. These 20 exams included 152 written questions, comprised of 116 essays, and 36 PT questions. Test results from first-time takers (more than 72,000) were included in the analysis. The analysis considered three primary variables: gender, race/ethnicity, and law school type. Secondary variables, such as item type, subject matter, and administration window were used to investigate trends within the gender, race/ethnicity, and law school type DIF results.

In comparison between male and female exam takers, the analysis identified moderate to large DIF in 20 percent of the questions evaluated, with female takers performing consistently better than their male counterparts. Gender-based DIF effect tended to be concentrated in less frequently selected topics, including trusts, wills, and community property.

With regard to race/ethnicity effects, DIF flags were identified in 16 percent of the questions for African American takers, with White takers as the reference group. The proportions of questions identified with DIF for Asian and Latino takers were lower, at 7 and 5 percent, respectively. All of the DIF results by race/ethnicity are in favor of White test takers. The analysis results for African Americans, however, were less reliable than the others identified because of the large differences in sample size between African American and White bar exam takers. When all nonwhite takers are grouped together in comparison with White takers, the overall DIF results were flagged for only 5 percent of questions reviewed.

The largest DIF results appeared across law school or applicant types, with California ABA law schools as the reference group. Applicants with a foreign JD had 75 percent of the questions flagged with DIF; about 65 percent for takers from registered law schools; 55 percent for out-of-state attorneys; and about 30 percent for California Accredited and out-of-state ABA law schools.

¹² See Goal 2 of the 2017-2022 State Bar Strategic Plan (Updated March 2020), <https://www.calbar.ca.gov/Portals/0/documents/bog/Updated-2017-2022-Strategic-Plan.pdf>

Considering the proportion of question items flagged with DIF, as well as the size of the DIF indicator, the overall results of the DIF study reported no major areas of concern for the CBX by gender and racial/ethnic groups. Nonetheless, to better understand the results of the DIF analysis and to proactively monitor for DIF in the future, the report recommended that the State Bar:

- conduct a retrospective bias and sensitivity review of the items flagged by the DIF study to inform future item development;
- incorporate proactive bias and sensitivity reviews of bar exam questions into the examination development process;
- use the results of the DIF study to inform the future design of the CBX with respect to item type selection;
- Review the CBX scoring process and any rubrics or methods for choosing, training, or assigning raters.

Review of the California Bar Examination Administration and Associated Components

Shortly after the disclosure of Bar Exam topics in July 2019, the State Bar requested that the California Department of Consumer Affairs (DCA), Division of Programs and Policy Review (DPPR), conduct a review of the administration of the CBX relative to best practices as defined by the Standards for Educational and Psychological Testing. The purpose of the review was to identify gaps and make recommendations for conforming more closely to the Standards. While the evaluation focused on operational and administrative aspects of the bar exam, standards related to exam validity issues were also reviewed.

To gather information on various aspects of bar exam administration, DCA staff reviewed operational manuals and procedures and conducted interviews with key Admissions staff over a period of several months in late 2019. Overall, the review concludes that the State Bar is meeting professional guidelines and technical standards in most areas.

For further improvement of bar exam operations however, the report recommended that the State Bar:

- formulate a policy on examination validation;
- conduct a practice analysis approximately every five years; and
- use subject-matter experts throughout validation activities with the understanding that appropriate subject-matter experts are:
 - currently licensed;
 - in good-standing;
 - representative of the diversity of practice areas;
 - representative of diverse geographic regions and experience; and
 - not members of the Board of Trustees, deans, or other law school educators, to avoid the appearance of conflicts of interest.

- engage in discussions about the most fair, reliable, and valid tool(s) to assess entry-level competence to practice as an attorney in California specifically considering:
 - reducing the essay portion of the CBX to standardize the process;
 - identifying the most critical content;
 - improving security issues associated with a two-day test; and
 - administering the MBE and the written portion separately and entirely by computer.
- improve standardization in the scoring of essay and performance questions with consistent rubrics, weights, and linkage to minimum competence;
- consider using vignette-style questions administered by computer to assess competency not captured in essay or performance task questions; and
- strengthen policies for responding to inappropriate test-taking behavior when it occurs.

A Report on the Effects of the CBX Phased Grading Process

The State Bar’s 2017 Governance in the Public Interest Task Force report included a review of the State Bar’s various volunteer subentities, including the CBE. Among the recommendations related to the CBE was a recommendation to review the bar exam grading process. Pursuant to this recommendation, State Bar staff consulted with the CBE’s psychometrician, Dr. Roger Bolus, to evaluate the processes for grading the CBX.

One important element of the grading process relates to the re-reading of exams that fall short of a passing score of 1440, but are above 1390, referred to as the “regrade band.” Currently exams that fall into the regrade band are read and scored again. In certain cases, exams may be read a third time.

Following a second read of exams, a new score is calculated based on the average scores of the first and second read. At that point, an average score of 1440 or above would receive a pass outcome while exams with average scores below 1440 follow one of two paths depending on the score difference between the first and second read. If the discrepancy between the two reads is smaller than 10 points for each of the questions, the initial fail outcome from Phase I becomes the final outcome. But if the discrepancy between the two reads is larger than 10 points on any written question, that exam is graded a third time.

To evaluate the costs and benefits of the regrading process, the study of phased grading looked at exam outcomes from the first four administrations of the two-day CBX, from July 2017 to February 2019. Out of 25,700 total test takers during that period, about 13 percent – almost 3,300 exams – went into the regrade process. It is estimated that the regrade process extended the release of bar exam results by 14 and 21 days on the February and July administrations, respectively.

In terms of outcomes for exam takers, the regrade process resulted in increasing the overall CBX passing rates by between 1.33 and 1.97 percent in the 4 administrations of the exam evaluated.

Out of about 3,300 exams that went into the regrading process, the third read changed the outcome for only 21 applicants. Thus, by itself, the third read added less than 0.1 percentage point to the overall pass rate. In addition, within the regrade band, the closer the initial score was to 1440, the greater the likelihood that regrading would result in a passing outcome. An analysis of the regrade outcome for each 5-point increment within the regrade band found that *no applicants* with initial scores from 1390 to 1399 received a passing outcome after regrade in any of the four exams reviewed.

In addition to grading time required and the impact on pass rate, the study evaluated other factors that might affect either the grading timeframe or exam outcomes, such as the number of graders and the average speed of grading exam papers. After taking into consideration all relevant factors and running through simulations of various models, the study recommended the following changes to achieve the dual goals of reducing the time to grade the bar exam while maintaining the integrity of the grading process:

- Eliminate the third read (Phase III) grading;
- Compress the regrade band for Phase II from 1390 – 1439.99 to 1400 – 1439.99; and
- Add two graders to each CBX question grading team.

The CBE has been involved to varying degrees in the four studies summarized in this agenda item. Current CBE Chair, Robert Brody, was appointed as a member of the CAPA Working Group. Mr. Brody provided updates on CAPA at each CBE meeting during the pendency of the study. The consultant to the CAPA study and State Bar staff to the study also provided periodic updates to, and sought the input of the CBE on the work.

Members of the CBE were intentionally shielded from the planning and execution of the DIF study. Early in the development of the DIF study the concern was raised that participation of members of the CBE in this effort might, inadvertently, influence the selection of questions on the CBX. To avoid this, the research team planned to bring the results of the study to the CBE and solicit their involvement only after the study was completed. Accordingly, the report on the DIF analysis was presented to the CBE at its April 24, 2020, meeting.

Regarding the review conducted by the Department of Consumer Affairs on the Administration of the bar exam, the CBE was told that the State Bar had contracted with the DCA to evaluate examination administration processes to ensure that the State Bar adhered to best practices for test administration and security, and the CBE was briefed on the project at the August, 2019 CBE meeting with report findings presented at the April 24, 2020, meeting.

The CBE was also provided with the results of the phased grading analysis at its April 24 meeting. The CBE took action on these reports where it was empowered to do so. Specifically, the CBE approved recommendations related to phased grading – the elimination of the Phase III read, the compression of the re-read band, and the addition of two more people to the grading teams.

As outlined in the proposed roadmap for short-term and long-term action, below, staff recommends that several of the items outlined in the recently completed studies that remain unaddressed be forwarded to the CBE for action.

A Roadmap For Bar Exam Development and Improvement

Recommendations for Short-Term Action

Each of the following recommendations, which tie directly back to one of the four reports summarized in the present agenda item, is organized within the life-cycle framework of a testing program and appears amenable to a relatively short-term implementation plan. As noted above, the CBE, under authority granted through the Board of Trustees has already approved the recommendations under “Administration and Scoring.”

Planning and Evaluation	Design and Development	Administration and Scoring
<p><i>Minimum Competence Definition:</i></p> <ol style="list-style-type: none"> 1. Adopt construct statement and entry-level attorney definition; 2. Further develop and operationalize minimum competence definition; and 	<p><i>Review Exam Questions:</i></p> <p>Convene a diverse panel, including members of the Council on Access and Fairness, to review flagged items from the DIF study and to develop guidelines to mitigating against DIF prospectively</p>	<p><i>Modify Phased Grading:</i></p> <ol style="list-style-type: none"> 1. Eliminate Phase III grading; 2. Narrow second read band; and 3. Add graders
	<p><i>Modify Grading Policies:</i></p> <p>Modify grading policies to develop model answers and scoring rubrics at the time of question development and pretest.</p>	<p><i>Diversity of Graders and Observers:</i></p> <ol style="list-style-type: none"> 1. Modify policies regarding graders to authorize entry-level attorneys to serve as graders and; 2. Ensure geographic and practice type diversity; 3. Evaluate grader compensation levels and modify policy regarding categories of permissible observers <ol style="list-style-type: none"> 1. Revisit Cheating Policy 2. Revisit policy on intervention in cases of flagrant cheating.

Long-Term Policy Questions

The results of the CAPA study, in conjunction with the concurrent parallel undertaking by the NCBE, suggest the need for consideration of significant policy issues, including a foundational question of whether or not California will continue to develop its own bar exam. This question, and other related matters outlined below, will require a longer-term, deliberative planning process. As reflected in the Recommendations section below, staff recommends that this process be effectuated by a newly formed Blue Ribbon Commission on the Future of the California Bar Exam.

Among the questions that this Commission would address are the following:

1. Upon review of the final results from NCBE's 2020 practice analysis: should California transition to the Uniform Bar Exam (UBE);
If not, then:
 - a. Should the recommendations of the CAPA Working Group regarding legal topics and competencies be implemented??
 - b. Should the bar exam format to test competencies recommended by CAPA be adopted?
 - c. Should the bar exam cut score be reevaluated?
 - d. What topics should be tested outside of the bar exam?
 - e. If the CAPA study recommendations were implemented, what would the content and format of an independent, professional responsibility exam include?

If, however, the UBE were adopted:

2. Should a supplementary, California-specific exam be developed, and what topics and skills would such an exam test?

FISCAL/PERSONNEL IMPACT

There will be a workload impact associated with staffing the Blue Ribbon Commission which can be absorbed by existing State Bar personnel; operating costs to fund meetings and any needed consulting support, are not expected to exceed \$50,000.

AMENDMENTS TO RULES OF THE STATE BAR

None

AMENDMENTS TO BOARD OF TRUSTEES POLICY MANUAL

None

STRATEGIC PLAN GOALS & OBJECTIVES

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

Objective: n. Conduct a California specific job analysis to determine the knowledge, skills, and abilities for entry level attorneys. Upon completion, conduct a new content validation study

Goal: 4. Support access to legal services for low- and moderate-income Californians and promote policies and programs to eliminate bias and promote an inclusive environment in the legal system and for the public it serves, and strive to achieve a statewide attorney population that reflects the rich demographics of the state's population.

Objective: i. No later than March 31, 2020, identify ways that diversity and inclusion principles can be institutionalized in Bar exam development and grading analyses with final proposals, including any formal guidelines or rule proposals, to be submitted to the Board by December 31, 2020.

RECOMMENDATIONS

Should the Board of Trustees concur in the proposed action, passage of the following resolution is recommended:

RESOLVED, that the Board of Trustees directs State Bar staff to:

Establish a Blue Ribbon Commission on the Future of California Bar Exam in partnership with the California Supreme Court. The recommended charge of this Commission, to be finalized in consultation with the Supreme Court, would include:

- Review of the results of the California Attorney Practice Analysis and the recommendations of the CAPA Working Group;
- Review of the results of the 2020 National Conference of Bar Examiners job analysis and any next steps articulated by that body as related to CBX, or MBE or UBE content or format;
- Development of recommendations for the California Supreme Court and the State Bar of California regarding:
 - Adoption of the Uniform Bar Exam (UBE):
 - Specifications for a California-specific supplementary bar exam should the UBE be recommended for adoption.
 - Revisions to the California Bar Exam if the UBE is not recommended for adoption:
 - Legal topics to be tested;
 - Skills to be tested;
 - Testing format; and
 - Specifications for a supplementary professional responsibility exam if that topic were eliminated from the CBX.
 - Bar exam cut score

and it is

FURTHER RESOLVED, that the Board of Trustees directs the Committee of Bar Examiners to:

- Review CAPA Working Group recommendations regarding the definitions of an entry-level attorney and minimum competence and:
 - Recommend that the Board approve and recommend to the California Supreme Court as recommended by the Working Group; or
 - Recommend that the Board approve and recommend to the California Supreme Court amended definitions.
- Review Bar Exam grading policies:
 - Revisit grader eligibility criteria and compensation levels as outlined in the 2020 DCA report; and
 - Examine the recommended timeline for the development and content of grading rubrics as outlined in the 2020 DCA report.
- Modify Exam administration policies:
 - Revisit flagrant cheating policy as outlined in the 2020 DCA report.
- Work with the Council on Access and Fairness to:
 - Convene a panel charged with:
 - Reviewing questions flagged for DIF in the 2020 differential item function analysis; and
 - Developing guidelines for minimizing the risk of future differential item functioning.

ATTACHMENT(S) LIST

- A.** DRAFT - The Practice of Law in California: Findings from the California Attorney Practice Analysis and Implications for the California Bar Exam
- B.** State Bar of California: Differential Item Functioning Analysis Report
- C.** Review of the California Bar Examination Administration and Associated Components
- D.** A Report on the Phased Grading of the California Bar Examination: A Profile of Recent Results and Modeling the Impact of Alternative Approaches

Attachment A



The State Bar of California

The Practice of Law in California: Findings from the California Attorney Practice Analysis and Implications for the California Bar Exam

**Final Report of the
California Attorney Practice Analysis
Working Group**

DRAFT

April 20, 2020

Executive Summary

In 2018 the State Bar established the California Attorney Practice Analysis (CAPA) Working Group, with members appointed by the Supreme Court, and charged it with overseeing and providing guidance on a practice analysis study, and using the results of the study as the basis for providing recommendations on exam content. Practice analysis refers to “the systematic collection of data describing the responsibilities required of a professional and the skills and knowledge needed to perform these responsibilities.” The purpose of the practice analysis study overseen by the CAPA Working Group was to gauge alignment between the content of the California Bar Exam (Bar Exam, Exam, or CBX) and the practice of law in California.

The Working Group held its kick-off meeting in October 2018 with the support of a grant from the AccessLex Institute. Scantron, a consulting firm, was hired to design and execute the study. The Working Group endorsed the collection of data on attorney practices along two principal dimensions:

- what attorneys do as reflected in daily tasks; and
- what knowledge attorneys use to perform those tasks.

A subgroup of the CAPA Working Group (the Practice Analysis Panel) was created to draft detailed task descriptions and to construct a taxonomy of legal topics. The resulting list of 13 competencies (consisting of 110 discrete tasks) and 19 legal topics (expanded into 114 subtopics) was used to collect data on attorney practice using two different, complementary, survey methods – a traditional survey of ratings based on recollection of experience, and an Experience Sampling Method capturing data on attorneys’ work in real time.

After analyzing and discussing the findings from the surveys as well as analyses provided by State Bar staff, the CAPA Working Group came to a consensus on several recommendations relating to the content of the Bar Exam.

Recommendation One: As the State Bar of California engages in further and continuous study of the Bar Exam, it is recommended that the State Bar adopt the construct statement utilized by the working group to define the general scope of the Bar Exam and utilize the “first three years of practice” as the definition of “entry-level.” The construct statement is as follows:

The California Bar Examination assesses legal knowledge, competency areas, and professional skills required for the entry-level practice of law and the effective, ethical representation of clients.

Recommendation Two: Based on study results indicating the most frequent legal topic areas for entry level attorneys and due to the need to reduce the current repetition in testing skills, it is recommended that the following eight legal topics be adopted for a new Bar Exam content outline:

- Administrative Law and Procedure
- Civil Procedure
- Constitutional Law
- Contracts
- Criminal Law and Procedure

- Evidence
- Real Property
- Torts

As noted, the recommendation reduces the number of legal topics from 13 to eight. Although the idea of the potential advantages of reducing the number of legal topics tested, in particular to de-emphasize memorization, was a consideration of the Working Group, the recommendation to reduce the number of legal topics tested is grounded on the results of the study. The empirical findings revealed that the varied practices of entry-level attorneys in California encompass a broad array of distinct legal topics, not all of which could possibly be tested. Based on the data, the Working Group identified a set of core legal topics that would provide a sound basis for testing the fundamental skills of issue spotting and legal analysis that would not be repetitive and that would be sufficient to assess minimum competency for the entry level practice of law across the broad range of practice areas identified in the survey.

Recommendation Three: Based on the survey data of the competencies and professional skills necessary for entry-level attorneys, the Working Group identified six broad areas as the most relevant competencies and recommends these be assessed by the Bar Exam:

- Drafting and writing;
- Research and investigation;
- Issue-spotting and fact-gathering;
- Counsel/advice;
- Litigation; and
- Communication and client relationship

The first five of these are specific competences that were highly related as essential to the practice of law in the survey; the sixth Communication and client relationship was constructed by the Working Group from three other competencies that were also highly rated (Establishing the client relationship, Maintaining the client relationship and Communication) and deemed by the Working Group to be interrelated. The Working Group concluded that assessing these core competencies shown by the data to be essential to the entry-level practice of law, along with testing the core group of legal topics, would ensure a more reliable, documented link between actual practice and bar exam content.

Introduction

In 2017 the State Bar completed two studies specifically related to the California Bar Examination (CBX or Exam): a standard setting study to evaluate whether the pass score was set appropriately to reflect the minimum competence of an entry-level attorney, and a content validation study to evaluate whether the topics covered on the Exam aligned with the knowledge, skills, and abilities required of entry-level attorneys.¹

While these studies represented important milestones in comprehensively assessing two critical elements of the CBX, the scope of the studies and the conclusions they reached were constrained by several factors. One of the most important constraints related to the content validation study which relied heavily on a national survey of practicing attorneys conducted by the National Conference of Bar Examiners (NCBE) in 2012. The national focus of the NCBE study – while appropriate for the NCBE’s purposes – lacked California-specific content for the evaluation of the written component of the CBX.

To remedy this, in 2018 the State Bar sought consulting support from specialists in the fields of psychometrics and practice analysis to conduct a study to describe the practice of law in California for entry level attorneys, evaluate the link between Exam content and current legal practice, and use that information to create a Blueprint – an outline of content coverage across legal topics and job responsibilities – for future Exam development and selection of Exam topics and question items.

The Board of Trustees of the State Bar created the California Attorney Practice Analysis (CAPA) Working Group to guide the consultant’s work and recommend legal topics and competencies to include on the CBX. The Supreme Court appointed members of the Working Group to ensure broad representation across the spectrum of attorney practice and demographics in California.

This Executive Report of the CAPA Working Group begins with a discussion of *practice analysis*, the centerpiece of the work overseen by the CAPA Working Group. It then describes the formation of the CAPA Working Group, its charge, and activities. The remainder of this Report describes the findings of the practice analysis. It explains recommendations of the CAPA Working Group regarding legal topics and job responsibilities to include on the Exam, and concludes by making recommendations for a process to move forward and incorporate the findings of the practice analysis into the future design, development, and administration of the CBX.

Practice Analysis

Practice analysis, sometimes referred to as job analysis, refers to “the systematic collection of data describing the responsibilities required of a professional and the skills and knowledge needed to perform these responsibilities.”² Practice analysis conducted for the purpose of licensure adds an additional dimension to this definition by focusing on the knowledge, skills, and abilities (KSAs) of people *at the entry level* of a profession.

¹ For a summary of these reports, see Report to the Supreme Court of the State of California Final Report on the 2017 California Bar Exam Studies, December 1, 2017, <http://www.calbar.ca.gov/Portals/0/documents/reports/2017-Final-Bar-Exam-Report.pdf>

² See “Practice Analysis: Building the Foundation for Validity,” Joan E. Knapp and Lenora G. Knapp, in *Licensure Testing: Purposes, Procedures, and Practices*, James C. Impara, editor, 1995.

Documentation of entry-level job KSAs is essential for all professions that use testing to confer licenses on new members because exams need to accurately assess the actual practice of the profession. Without this type of analysis, there is no connection between the content of an exam and the practice of the licensees. Moreover, these studies need to be updated periodically as changes over time in the practice of a given profession could easily drift away from the content of the licensing exam.³

While practice analysis is common to all licensing exams, the specifics of each profession create unique challenges for documenting actual practice. For the legal profession in California, the domains of law, practice settings, and tasks performed by attorneys across the state are vast. And, distinguishing between the totality of work that attorneys perform and the work that an entry-level attorney should know requires careful analysis.

Practice analysis relies heavily on the empirical study of actual practitioners. The only certain way comprehensively to capture the broad scope of work performed by attorneys for the purpose of understanding the knowledge, skills and abilities necessary to the practice of law is to ask them in a methodologically rigorous manner. In the summer of 2018 the State Bar issued a Request for Proposals, seeking consulting services from a vendor with the technical ability and capacity to conduct just such a study of attorney practice.

In the process of evaluating the vendors, the State Bar considered not only vendor experience in conducting practice analysis studies in a broad range of professions but also their flexibility in considering innovative methods. The State Bar submitted a grant application to the AccessLex Institute to fund the study, with the key element of the application centered on a proposal to collect real-time data on attorney practice using a survey method called Experience Sampling Method (ESM). The AccessLex Institute awarded the grant in the summer of 2018, around the same time when Scantron was selected as the vendor due to its expertise on a full range of psychometrics services, as well as their enthusiasm in the cooperative effort to test the new ESM approach along with more established traditional method.

The CAPA Working Group

In the fall of 2018 the State Bar sought nominations for membership in the CAPA Working Group. The Supreme Court then appointed a group of attorneys from the nominees who were broadly representative of the legal profession and key stakeholders to serve on the Working Group. CAPA Working Group members included attorneys from most geographic regions of the state, those who work in private and public practice, attorneys in large and small firms, Deans from California law schools, two judges, a representative of the California State Legislature and an out-of-state representative who previously served as the Chair of the Board of the National Conference of Bar Examiners. The Chair of the State Bar Board of Trustees was appointed to serve as the Chair of the CAPA Working Group. (See Appendix A.)

³ A recent amendment by the Supreme Court to California Rule of Court 9.6(b) , understanding the need to periodically update practice analyses, mandates the State Bar “conduct an analysis of the validity of the bar examination at least once every seven years.” Validity is commonly used in the social sciences to refer to the linkage between *a concept* and *the measures* used to describe and evaluate that concept. In the case of the CBX, practice analysis is needed to assess how accurately the questions on the exam – the measures – capture the KSAs necessary to practice law in California at an entry level of proficiency – the concept.

At the Working Group's first meeting, Scantron recommended the creation of a smaller, more focused group to further develop the broad conceptual categories and lists of legal topics and attorney responsibilities that would be used to collect survey data on attorney practice. A Practice Analysis Panel (Panel) was created in January, 2019 with membership designed, as with the Working Group as a whole, to reflect a broadly diverse attorney population. (See Appendix B.)

During a three day meeting in March 2019, the State Bar and Scantron facilitated large and small group discussions with the Panel. An instructional booklet was provided to Panel members with a description of the target audience for the CBX and definitions of key terms that needed to be categorized in order to capture the data on attorney practice in a survey. The Panel agreed upon the following structure for the survey design:

- *Areas of responsibility* are the major responsibilities or duties that comprise the practice of law. Included within areas of responsibility are *competencies*, which are more focused responsibilities that may be carried out within an area of responsibility;
- A *task statement* defines an activity that elaborates on an area of responsibility. The set of task statements for a particular area of responsibility offers a comprehensive and detailed description of that responsibility;
- *Legal topics* are major subject areas in law. They may be augmented with more specific topics, or *subtopics*, that offer detail about the broader legal topics.⁴

The outline for the survey created by the Panel establishes four broad areas of attorney responsibility, 13 competencies associated with those responsibilities, and 110 tasks associated with those competencies. The outline also identifies 19 legal topics and 114 subtopics. Scantron conducted a crosswalk analysis comparing this outline with similar outlines developed by the National Conference of Bar Examiners (NCBE) and other organizations to ensure that the categories of legal topics and attorney responsibilities used by the CAPA comprehensively captured the practice of law in California. The crosswalk analysis confirmed the comprehensive coverage of the CAPA survey.

The full CAPA Working Group reviewed the outline and the crosswalk analysis in April 2019 and approved these to serve as the foundation for data collection. Table 1 shows the areas of responsibilities and their corresponding competencies and Table 2 shows the legal topics recommended by the Panel. The subtasks associated with each competency and the subtopics associated with each legal topic are provided in Appendix C

⁴ In its technical report, Scantron refers to responsibilities and legal topics as performance and content domains, respectively.

Table 1: Areas of Responsibility and Competencies

Area of Responsibility	Competencies
Establishing and Maintaining Relationships	<ul style="list-style-type: none"> • Establishing the client relationship • Maintaining the client relationship • Communicating with others
Practice Management and Administration	<ul style="list-style-type: none"> • Practice management • Case or matter management • Supervision and collaboration
Factual and Legal Analysis	<ul style="list-style-type: none"> • Issue spotting and fact gathering • Research and investigation • Drafting and writing
Resolutions	<ul style="list-style-type: none"> • Counsel/advice (via conversation, in-person, or by telephone) • Negotiation and closing • Alternative Dispute Resolution • Litigation

Table 2. Legal Topics

Administrative Law and Procedure	Family Law
Business Associations	Finance (includes bankruptcy)
Civil Procedure	Legislation
Constitutional Law	Professional Responsibility
Contracts	Real Property
Criminal Law and Procedure	Secured Transactions
Employment	Securities
Environment and Land Use	Tax
Estate Planning, Trusts, and Probate	Torts
Evidence	

Survey Design, Administration, and Analytical Decisions

Once the categories for data collection were established, the Working Group focused on the development of a survey to capture empirical data on the practice of law in California. To serve the purpose of the CAPA Working Group, the survey needed to capture a number of different pieces of information about areas of responsibility and legal topics including:

- the *frequency* with which attorneys practice in specific areas of responsibility and legal topics;
- the *criticality* of proficiency in tasks and legal topics in attorney practice;
- the *performance expectation*, that is, whether the performance of a particular task was expected of entry-level attorneys or was considered appropriate only to more experienced attorneys.

The CAPA Working Group also evaluated the relative merits of conducting a traditional practice analysis survey (Traditional Survey) and conducting a survey using the Experiential Sampling Method (ESM) and concluded that both would be used. In traditional practice analysis surveys, respondents are asked to recall their experience working in different legal domains and on different tasks during the previous 12 months. The respondents are asked to rate the frequency with which they worked in the different areas and tasks and, for the CAPA, would also be asked to rate the criticality of the work. Given the broad scope of attorney practice that needs to be covered, a survey like this typically takes at least 40 minutes to an hour to complete.

ESM, in contrast, is based on short, web-based surveys delivered by email at random times during the work day. Rather than capture information that attorneys recall from the previous 12 months, ESM asks attorneys to report what they are working on in the moment that they receive the survey. The speed with which ESM surveys can be completed – no more than 2-3 minutes – allows for attorneys to respond to multiple surveys per day. The aggregation of that data, collected from a large sample of attorneys, creates a large, statistically robust sample of detailed data on attorney practice.

In addition to the benefit of capturing data on attorney practice in real time, there are further methodological benefits of the ESM. This method also allows for the collection of more detailed data related to legal domains and responsibilities because the respondent is only answering questions related to the work that is taking place at that particular moment. While it would be unrealistic to expect attorneys to recall all of the tasks that they performed over the previous year, the ESM allows for a respondent to provide information directly related to the tasks they were performing at the time of responding to the survey. So, for example, attorneys responding to the ESM who indicated that they were engaged in the competency of “Drafting and Writing” were then given further options to select such as “Drafting pleadings, motions, statements, or briefs” and “Drafting correspondence.”

Both the Traditional and ESM Surveys asked respondents to rate the criticality and performance expectations of the competencies and legal topics associated with their legal work using a five-point scale. And both surveys used branching logic to direct respondents only to rating scales that applied to the competencies and legal topics that they worked on. The ESM survey differed slightly from the traditional survey in that it captures a *frequency* of task performance and usage of legal topic based on actual responses: the observed distribution of responses can be used as the actual distribution without needing attorneys to estimate the frequency with which they work in particular competencies or on specific legal topics. The ESM also included an additional scale to measure depth of knowledge (DOK) of legal topics. As a measure of cognitive complexity involved in applying knowledge, ranging from simple recall of factual information to higher level analysis and synthesis of more complex concepts, the DOK data provides another dimension related to the work of entry-level attorneys. For a summary of each rating scale see Table 3.

Pilot surveys based on the initial categories approved by the Working Group were administered to over 500 practicing attorneys in June 2019 for the purpose of assessing survey functionality and clarity. Improvements and modifications to the surveys were made based on this beta testing.

Table 3. Rating Scales

Construct	Traditional Survey	Experience Survey Method
<p>Frequency: How frequently the attorney performed the work activity or applied the legal topic in their practice</p> <ul style="list-style-type: none"> ✓ competencies ✓ legal topics 	<p>“On average, how frequently did you perform this work activity or apply the legal topic in your practice during the past 12 months?”</p> <p>1 = Never 2 = Rarely (once per year) 3 = Sometimes (once per month) 4 = Often (once per week) 5 = Repeatedly (daily)</p>	<p>Frequency was computed as the total number of times respondents report performing competencies or using legal topics</p>
<p>Criticality: The degree of harm (legal, financial, psychological, or emotional) that may result for clients and/or the general public if an attorney is not proficient.</p> <ul style="list-style-type: none"> ✓ competencies ✓ legal topics 	<p>“What degree of harm may result to clients and/or the general public if an attorney is not proficient in [competencies and legal topics]?”</p> <hr/> <p>1 = No harm at all 2 = Minimal harm 3 = Moderate harm 4 = Substantial harm 5 = Extreme harm</p>	<p>“What degree of harm may result if you are not proficient in performing the task related to [competency]”</p>
<p>Performance Expectations: Point in legal careers attorneys were first expected to perform the competency.</p> <ul style="list-style-type: none"> ✓ competencies 	<p>“When were you first expected to perform this task?”</p> <p>1 = 0 to 6 months of practice 2 = 7 to 12 months of practice 3 = Years 2 to 3 of practice 4 = Years 4 to 5 of practice 5 = After the fifth year of practice</p>	<p>“At what point in your legal career were you expected to first perform this task?”</p> <p>1 = First year 2 = Second year 3 = Third year 4 = After third year</p>
<p>Depth of Knowledge: Cognitive activity performed that demonstrates level of knowledge required when performing the competency.</p> <ul style="list-style-type: none"> ✓ legal topics 	<p>N/A</p>	<p>“What level of knowledge of [legal topic] is required when performing the task?”</p> <p>1 = Recall from memory 2 = Understand 3 = Apply 4 = Analyze 5 = Synthesize/Evaluate</p>

Sampling

In addition to the Traditional and ESM Survey formats being administered pursuant to CAPA, the project sampling strategy accounted for a concurrent practice analysis being conducted by the National Conference that the State Bar had agreed to provide the necessary support and collaboration. Because there are over 190,000 active, licensed attorneys in California, the CAPA Working Group determined that it did not need to choose between the two survey methods. Instead, the practice analysis could use both methods to capture data on the legal profession. This would allow for the Working Group to realize the benefits of each method and even provide for cross-validation of findings. Prior to selecting samples of attorneys to participate in one of the two surveys, however, the CAPA Working Group was notified that the National Committee of Bar Examiners was conducting its own practice analysis survey and was seeking participants from California, opening the opportunity for further cross-validation.

To accomplish the goal of administering all three surveys in a manner that would allow for cross-validation of results, the State Bar divided the pool of active, licensed California attorneys into three groups of approximately 63,000 each. Potential survey participants were randomly assigned to one of the three groups – one for the NCBE survey, another for the Traditional Survey, and a third to the ESM Survey.

Responses to the two CAPA Working Group Surveys were originally scheduled to be completed by August 26, 2019. To improve the survey response rate, that deadline was extended for a week, through September 3, 2019. The response rate for the traditional survey was 8 percent with 5,100 respondents, while the ESM survey's response rate was higher at 18 percent (11,090 respondents). However, because participants in the ESM survey had the opportunity to respond to the survey up to three times a day during their assigned survey week, the ESM survey yielded over 74,000 responses on attorneys' daily activities.

Synthesizing Data from the Traditional and ESM Surveys

The CAPA Working Group held three in person meetings and a conference call in the fall of 2019 to review the survey results and formulate their recommendations for a California Bar Exam Blueprint. Scantron prepared all results so that each survey's results could be compared against one other. Detailed findings of all of the results reviewed by the Working Group are available in the Scantron report.

The results generated extensive discussion which resulted in the Working Group making key decisions regarding the most effective way to review the findings, especially when results across the surveys were inconsistent.

- First, the Working Group decided to rely upon criticality ratings from the Traditional Survey in their evaluation of the importance attorneys attached to their work - whether in performing a task or using certain knowledge. This decision was based on the Working Group's assessment that the Traditional Survey's format allowed respondents to evaluate survey items within the larger context of an attorney's work. The Working Group determined that this context was valuable and, ultimately, a more accurate assessment of criticality, even if it might overestimate the criticality of the legal matters being handled;
- Second, the Working Group concluded that the ESM survey's estimates of the frequency of competencies and legal topics used by entry-level attorneys in their practice were more

accurate than those generated by the Traditional Survey. Indeed, the narrow range of the five-point scale used in the Traditional Survey provides very little variation to assess differences in frequency of competencies and legal topics. In contrast, the ESM survey shows the actual variance in frequencies to allow for a more reliable assessment of often quite substantial differences across categories;

- Finally, the Working Group agreed to combine data from the two surveys into standardized composite measures for the purpose evaluating and selecting competencies and legal topics for the CBX content outline.⁵ To combine the results from both surveys, the data were transformed to equivalent measures in which the mean value of each survey was set to 10, and one standard deviation was set to 5. The frequency and criticality data were transformed to a standardized scale first, and a single composite score was then created by averaging the two scores. Based on this single composite score, an item with a high score indicates that entry-level attorneys not only considered it important but performed or used it frequently.⁶

Findings from the CAPA Surveys

The following section describes the findings from the CAPA surveys and the Working Group’s evaluation of those findings using the scales described above to answer the following questions:

- What defines an “entry-level” attorney?
- What are the competencies expected of entry-level attorneys and how critical is proficiency in these competencies to the successful practice of law?
- What legal knowledge is required to perform those competencies?⁷

“Entry-level Attorneys” – Practitioners with Fewer than Three Years’ Experience

While the bar examination is designed to ensure the minimum competence of entry-level attorneys, some determination needs to be made as to *how long* an attorney can practice and still be considered “entry level.” In practice analysis studies, practitioners with fewer than three years of experience are generally considered “entry level” and data collection focuses on this group.

The CAPA study invited all active attorneys to participate in the survey to evaluate empirically the question of entry-level attorney practice. By collecting data on samples of all active attorneys, the CAPA surveys allowed for the evaluation of response patterns in relation to years of practice. All attorneys who participated in the study were asked when in their careers they were expected to be able to perform specific tasks. With data on the actual years of practice of the respondent attorneys, the CAPA study was able to compare these responses to practice data reported by attorneys at different points in their careers.

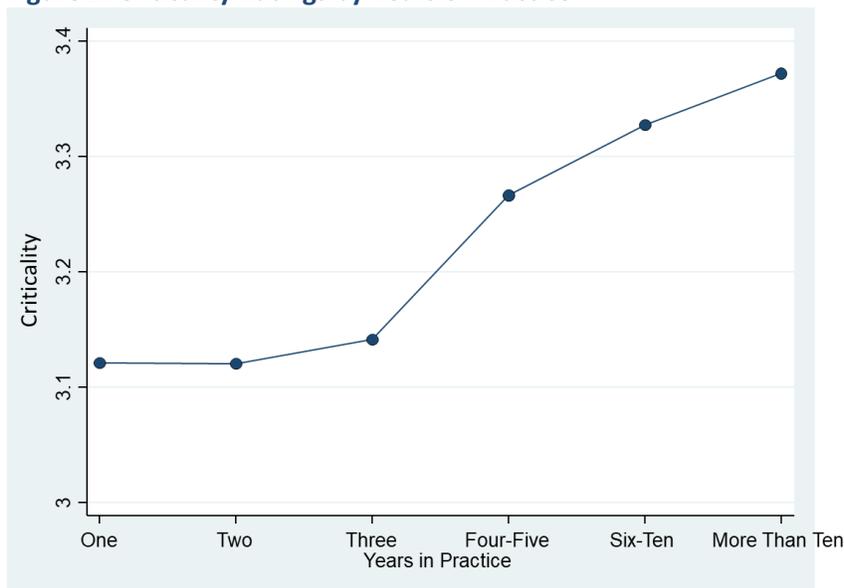
⁵ Data from the two surveys were weighted equally in the composite measure.

⁶ The calculation of the composite scores is explained in more detail in the Scantron report.

⁷ Although data were collected at the level of 110 tasks shown in Appendix A to document the practice of law in California, for purposes of developing an exam outline – a Blueprint – it is sufficient to analyze the data at the level of the broader categories of 13 competencies and 19 legal topics. The more granular task and subtopic descriptions will be useful at a later stage when developing detailed test specifications.

Figure 1 shows the relationship between attorneys' years of practice and their assessment of the *criticality* of their work. Criticality is defined as the potential harm to the client if the attorney lacks the competence to perform the specified task. Attorneys in their first three years of practice rated the criticality of their tasks at similar levels. As attorneys gained 4 to 5 years of experience, the criticality ratings began to rise at small but continual increments.⁸

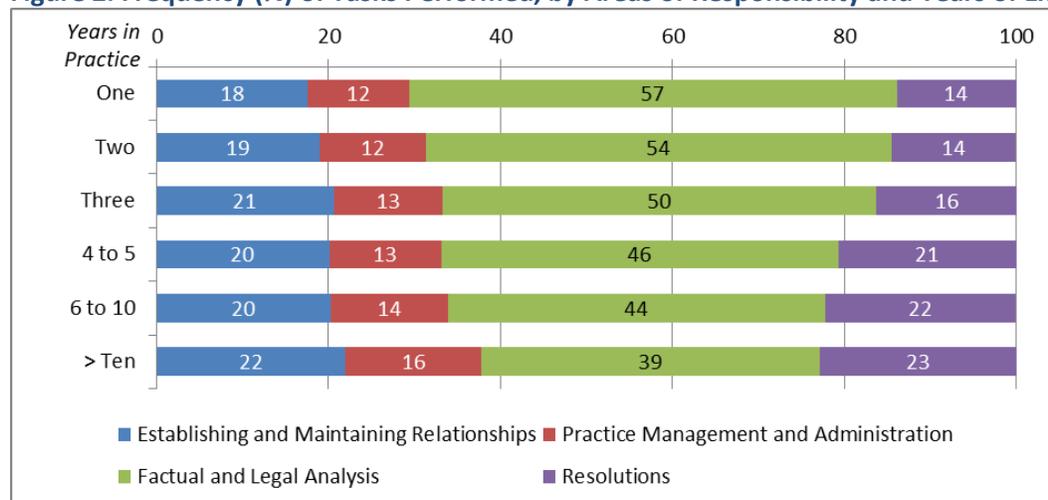
Figure 1. Criticality Ratings by Years of Practice



Analysis of the *tasks* performed by attorneys reveals gradual shifts of responsibilities as attorneys gain more experience. Figure 2 shows that tasks related to Factual and Legal Analysis account for a larger proportion of the time of newer attorneys – half or more of the time of attorneys in their first through third year of practice. The amount of time devoted to Factual and Legal Analysis and only declines to less than 50 percent in the fourth year of practice. As the amount of time devoted to Factual and Legal Analysis declines, the amount of time spent on Resolution, Establishing Relationships, and Practice Management and Administration increases.

⁸ Findings from the ESM survey.

Figure 2. Frequency (%) of Tasks Performed, by Areas of Responsibility and Years of Experience



Recommendation One: Language and Definitions for Future Bar Exam Studies

The findings regarding criticality ratings and frequency of tasks performed by new attorneys supports the use of the first three years to define the experience of an entry level attorney. The data show that the competencies expected of attorneys in their third year are very similar to the competencies expected in their first year, lending support to the use of a three year period for the definition of entry-level. The results reported below focus on entry-level attorneys.⁹

After delineating empirically the definition of entry-level attorney within the first three years of practice, the Working Group developed a construct statement for the bar exam as a framework for further defining the minimum competency for entry-level attorneys, as following:

The California Bar Examination assesses legal knowledge, competency areas, and professional skills required for the entry-level practice of law and the effective, ethical representation of clients.

The working group’s creation of the construct statement, followed by the deliberation and decision on the legal topics and competencies that are considered important for the bar exam, was instrumental in defining minimum competency for entry-level attorneys. At a more detailed level the term minimum competency needs further refinement so that it can be operationalized for different aspects of the bar exam such as the grading rubric and standard setting.

Toward a California Bar Exam Blueprint

While the findings from the surveys were critically important to the final deliberations of the Working Group, the scores alone, were not dispositive. In addition to the data from the CAPA surveys, the Working Group considered other relevant factors including whether the content might be better

⁹ For complete survey results, see the Scantron technical report.

delivered and assessed as part of a new attorney’s mandatory continuing legal education, or whether a topic was foundational to an understanding of the law, itself, even if not directly, frequently used.

Consideration of these additional factors led to the exclusion of some topics that the data, alone, might have indicated belong on the Blueprint. Conversely, these same value considerations led to the inclusion of topics that the data, alone, might not have suggested belong on the Blueprint.

Recommendation Two: Legal Topics

Table 4 shows the results of the data analysis of legal topics. These results are a composite of the results of the Traditional and ESM surveys, described above. Using Civil Procedure as an illustration of how the composite score works, a review of Table 4 shows that entry-level attorneys reported a frequency using Civil Procedure almost three standard deviations above the mean frequency for legal topics on the ESM survey (recalling that the composite scores set each of the individual scores to a mean of 10 with a standard deviation of 5), and entry-level attorneys rated Civil Procedure at almost one and a half standard deviations above the mean criticality ranking on the Traditional Survey.

The average of these two scores, weighted equally, gives Civil Procedure a composite score of 20.7. Compared to an item with a composite score near the mean value of 10, such as Constitutional Law at 9.6, Civil Procedure’s composite ranking is higher by more than 2 standard deviations, an indication of significant difference between the two legal topics as far as entry-level attorneys’ practice is concerned both in frequency and criticality.

Table 4. Standardized and Composite Ratings for Legal Topics

Legal Topic	Standardized Frequency – ESM Survey	Standardized Criticality – Traditional Survey	Composite
Civil Procedure	24.1	17.3	20.7
Professional Responsibility	9.6	19.4	14.5
Torts	13.3	15.2	14.2
Contracts	17.1	11.0	14.1
Evidence	10.0	17.3	13.6
Employment	15.4	11.0	13.2
Criminal Law and Procedure	11.7	13.1	12.4
Family Law	9.8	13.1	11.4
Administrative Law and Procedure	11.8	8.9	10.4
Constitutional Law	6.1	13.1	9.6
Estate Planning, Trusts, and Probate	10.0	6.8	8.4
Real Property	10.1	4.7	7.4
Environment and Land Use	5.6	8.9	7.2
Legislation	8.2	4.7	6.5
Business Associations	7.8	4.7	6.3
Tax	5.6	6.8	6.2
Finance (includes bankruptcy)	5.2	6.8	6.0

Securities	4.7	4.7	4.7
Secured Transactions	3.8	2.6	3.2

Note: Sorted in descending order by the composite rating.

The Working Group met several times to consider and discuss these empirical findings in the context of their collective understanding of the legal profession. All members of the Working Group were in agreement that the legal topics currently tested on the Bar Exam are important, as are the legal topics not currently tested but included in the two surveys. However, the charge of the Working Group was to provide recommendations as to content for the California Bar Exam, and the Working Group determined that fewer topics, not additional ones, should be tested specifically on the Bar Exam.

Although the idea of the potential advantages of reducing the number of legal topics tested, in particular to de-emphasize memorization, was a consideration of the Working Group, the recommendation to reduce the number of legal topics tested is grounded on the results of the study. The empirical findings revealed that the varied practices of entry-level attorneys in California encompass a broad array of distinct legal topics, not all of which could possibly be tested. Based on the data, the Working Group identified a set of core legal topics that would provide a sound basis for testing the fundamental skills of issue spotting and legal analysis, not be repetitive, and be sufficient to assess minimum competency for the entry level practice of law across the broad range of practice areas identified in the survey.

The initial threshold for inclusion in the recommended topics for testing was a composite score of 10 or higher. Five of the eight topics recommended both met that threshold and are already currently tested. After much discussion and debate, the consensus of the Working Group was to include two topics currently tested that did not meet the threshold and one final legal topic area (Administrative Law and Procedure) that did meet the composite threshold but is not currently tested. Specifically, the Working Group made determinations as to the specific topics to recommend for testing or removal.

The Working Group's recommendation is to INCLUDE:

- **Civil Procedure, Contracts, Criminal Law & Procedure, Evidence and Torts** – These five legal topics met the threshold composite score and are already being currently tested on the Bar Exam. In addition, these areas were affirmed as foundational by the Working Group.
- **Constitutional Law and Real Property** – While these topics ranked below the average of 10 on the composite scale, the Working Group agreed that Constitutional Law and Property are foundational and core knowledge for an understanding of the American system and common law. While fewer practitioners may specifically practice in these areas, the Working Group agreed that knowledge of constitutional and property rights is essential to the entry-level practice of law and the effective, ethical representation of clients.
- **Administrative Law and Procedure** – This topic met the threshold composite score. Administrative Law and Procedure covers a broad and diverse area of practice, encompassing procedures and regulations in professional licensing, unemployment insurance benefits, social security and retirement benefits, discrimination, labor and employment, workers' compensation, immigration appeals, securities, zoning, and many other regulatory issues. Administrative Law and Procedure also defines the scope of judicial review and establishes uniformity in the rulemaking process. Given these characteristics, the Working Group agreed

that Administrative Law and Procedure is a topic that is both foundational and practical, and thus should be tested on the bar exam.

Despite these topics meeting the composite score threshold, the Working Group's recommendation is to EXCLUDE:

- **Professional Responsibility** – While the Working Group acknowledged that ethics and professional duties to clients are as foundational as other topics indicated above, the Working Group agreed that the need for new attorneys to be competent in the area of ethics could be better addressed outside of the Bar Exam format. Currently all State Bar applicants must receive a passing score on the Multi-State Professional Responsibility Exam (MPRE) prior to licensure, and all State Bar applicants must have earned a passing grade in at least one course in ethics that addresses the model rules of professional conduct while in law school. In addition, the Working Group is aware of the New Attorney Training Program adopted by the State Bar in early 2018, which requires four hours of legal ethics training for new attorneys during the first year of their practice. The MPRE and course requirement, already in place, are more focused and dedicated assessments of an entry level attorney's knowledge of ethics and rules of professional conduct.
- **Employment Law, Family Law/Community Property and Business Associations** – The Working Group had a lengthy discussion about the inclusion or exclusion of these legal topics. As indicated by the composite score, many practitioners are engaged in aspects of employment law and family law, and while the score was lower for Business Associations, we know many entry level attorneys are working in and for business law firms. The Working Group determined however, that these areas are also largely a specialized practice and recommends alternatives to testing these topics on the Bar Exam including the possibility of specialized licenses, CLE courses, or coursework while in law school. The Working Group recommends further evaluation to determine alternative measures to ensure minimum competency in these areas.
- **Remedies, Trusts & Wills** – Based on the lower composite scores and the fact that concepts within Remedies and Trusts & Wills frequently blend with Contracts, Torts and other legal topics, the Working Group determined these topics could be removed from the testing topics.

In sum, the Working Group reached a consensus on recommending eight legal topics for the Blueprint, shown in Table 5. Compared to the legal topics included in the current CBX scope, also shown in Table 5, the recommended Blueprint for legal topics reduces the total number of topics from 13 to 8. Seven of the eight topics recommended by the Working Group are already included among the 13 current topics; one new legal topic is added – Administrative Law and Procedure; and six of the current legal topics on the CBX are recommended to be removed.

Table 5. Legal Topics on California Bar Exam and Recommended by CAPA Working Group

13 Legal Topics Tested on Current Bar Exam	8 Legal Topics Recommended by CAPA Working Group	Result of Recommendation
Civil Procedure	Civil Procedure	Retain
Constitutional Law	Constitutional Law	Retain
Contracts	Contracts	Retain
Criminal Law and Procedure	Criminal Law and Procedure	Retain
Evidence	Evidence	Retain
Real Property	Real Property	Retain
Torts	Torts	Retain
	Administrative Law and Procedure	Add
Business Associations		Remove
Community Property		Remove
Professional Responsibility		Test via MPRE, mandated course in law school, Mandatory legal education in year 1 after being licensed
Remedies		Remove
Trusts		Remove
Wills and Succession		Remove

Recommendation Three: Competencies and Professional Skills

Following the review of legal topics, the Working Group evaluated the composite scores for competencies shown in Table 6.

Unlike the decision making process for legal topics which required both the evaluation of composite scores and various policy and other considerations, the composite scores for competencies were deemed sufficient to guide the decision process in selecting relevant competencies expected of entry-level attorneys. There were no competencies that were rated low on the composite score that were considered essential to the practice of law, nor were there any highly ranked competencies that the Working Group determined might be amenable to treatment outside of the CBX.

There were, however, some competencies that were closely related to one another functionally, that the Working Group determined could be grouped together. Thus, the Working Group determined that Communication, Establishing Client Relationship, and Maintaining Client Relationship should be bundled together into a single, broader category. It is also recognized that, at least using the traditional testing method, some competencies are more amenable to traditional forms of testing than others. The Working Group discussed at length the challenges of testing these competencies using the traditional Bar Exam format and testing methods. Drafting and Writing, Research and Investigation, Issue Spotting and Fact Gathering have been tested using the current Performance Test. However, the Working Group agreed that new testing formats would be needed to more effectively test these competencies, in particular, Communication and Client Relationships.

Table 6. Standardized and Composite Ratings for Competencies

Competency	Standardized Frequency – ESM Survey	Standardized Criticality – Traditional Survey	Composite
Drafting and writing	24.0	15.5	19.7
Research and investigation	14.6	12.8	14.0
Litigation	9.7	16.8	13.1
Issue-spotting and fact-gathering	10.0	15.8	12.8
Communicating	10.7	11.2	11.1
Counsel/advice	7.5	13.2	10.3
Maintaining client relationship	8.3	11.1	9.8
Negotiation and closing	6.5	9.9	8.4
Case or matter management	8.4	7.6	8.1
Establishing client relationship	6.7	6.7	6.7
Practice management	6.2	3.0	4.6
Supervision and collaboration	7.4	0.9	4.0

Note: Sorted in descending order by the composite rating.

The final list of competencies recommended by the Working Group for consideration in the CBX blueprint is as following:

- Drafting and writing
- Research and investigation
- Issue-spotting and fact-gathering
- Counsel/advice
- Litigation
- Communication and client relationship including:
 - Establishing the client relationship
 - Maintaining the client relationship
 - Communication

How the CAPA results compare with the recent NCBE study

The National Conference of Bar Examiners (NCBE) launched its practice analysis study in 2018, about the same time as the launch of the CAPA study. Data collection for its survey was also completed in the fall of 2019. The two surveys share many similarities in both survey design and findings.¹⁰ There are also

¹⁰ See NCBE report: https://testingtaskforce.org/wp-content/uploads/2020/03/TestingTaskForce_Phase_2_Report_031020.pdf.

important differences, especially with regard to project scope and how the survey results are analyzed and synthesized to inform next steps.

Survey Design

NCBE's survey invited participation from all US licensed attorneys that fall into one of the following two categories:

- Newly licensed lawyers (termed "NLLs" in the report) who have been licensed for three years or less;
- Experienced attorneys who have had direct experience supervising NLLs.

Out of more than 30,000 participants who accessed the survey, the survey received valid responses from 3,153 NLLs and 11,693 non-NLLs, for a total of 14,846 responses.

The survey questions were phrased slightly differently for the two different groups. NLLs were asked to provide their ratings (frequency, criticality, etc.) on various questions according to their own practice experience. Experienced attorneys were asked to provide their assessments based on their experience directly supervising NLLs. The purpose of this design is to obtain two sets of responses on the same questions, allowing for comparisons of the responses between NLL's self-assessment in relation to what might be deemed more experienced assessment from non-NLLs. The survey findings are presented in the report with both responses side by side, most of time showing highly correlated results.

Organizing Framework of Survey and Comparability of Findings

As described above, the CAPA survey questions were organized into two major categories: tasks that the attorneys perform and knowledge and skills required to perform those tasks. NCBE's survey consists of four dimensions, as organized in the following categories:

- 179 tasks, grouped into 4 categories;
- 77 items for knowledge areas, without another layer of classification;
- 36 items for skills, abilities, and other characteristics (termed SAOs); and
- 24 items for technology.

The different organizing framework used in the two studies naturally produced survey questions that vary in specificity and comparability in the description of tasks and knowledge areas. For example, one task under the Competency area of Research and Investigation in the CAPA study is "Research laws and precedents." In NCBE's survey, in contrast, there are several task statements that would fit this single CAPA task, but given more specificity as to the different research activities, including:

- Research case law;
- Research statutory and constitutional authority;
- Research secondary authorities; and
- Research administrative regulation, rules, and decisional law

Table 7 shows the top ten rated tasks from the two studies. Ranking for the CAPA study was derived from the composite scores derived from the Traditional Survey's criticality rankings and the ESM

Survey's percentage frequency distribution. NCBE's ranking was based on a metric of percent performed. Despite the different granularity in describing the tasks and the different metrics in how they are measured, the two lists of ten items share the relatively high ratings given to a cluster of research-related activities for entry-level attorneys.

Table 7. Top Ten Rated Tasks from CAPA and NCBE Practice Analysis Study

CAPA Composite (ESM Frequency and Traditional Criticality)	Ranking
Draft pleadings, motions, statements, or briefs.	1
Research laws and precedents.	2
Review the documents collected.	3
Review relevant records and documents.	4
Identify legal and factual issues.	5
Edit drafts or documents.	6
Prepare for trial (e.g., subpoenas, exhibits, motions in limine, jury instructions).	7
Advise the client regarding the benefits, risks, and consequences of a course of action.	8
Calendar deadlines.	9
Communicate with opposing counsel or parties.	10
NCBE (Percent Performed)	
Identify issues in client matter including legal, factual, or evidentiary issues.	1
Research case law.	2
Interpret laws, rulings, and regulations for client.	3
Research statutory and constitutional authority.	4
Evaluate strengths and weaknesses of client matter.	5
Evaluate how legal document could be construed.	6
Develop specific goals and plans to prioritize, organize, and accomplish work activities.	7
Conduct factual investigation to obtain information related to client matter.	8
Research secondary authorities.	9
Consult with colleagues or third parties regarding client matters.	10

Table 8 looks at the treatment of legal topics: 77 items in the NCBE study, referred to as “knowledge area” compared to 19 legal topics evaluated in the CAPA study. NCBE's ranking was based on the average importance rating included in its report while the CAPA ranking was derived from the same composite score described above and used in the Working Group deliberations. It should be noted that gaps in the NCBE rank order, for example rankings from 5 to 9, result from knowledge areas in the NCBE survey that do not have direct, corresponding items in the CAPA list for legal topics.¹¹ A few items referring to the same subject matter but using different terminology were matched to facilitate the comparison. Criminal Law and Procedure was treated as a single item in CAPA but listed separately in the NCBE survey.

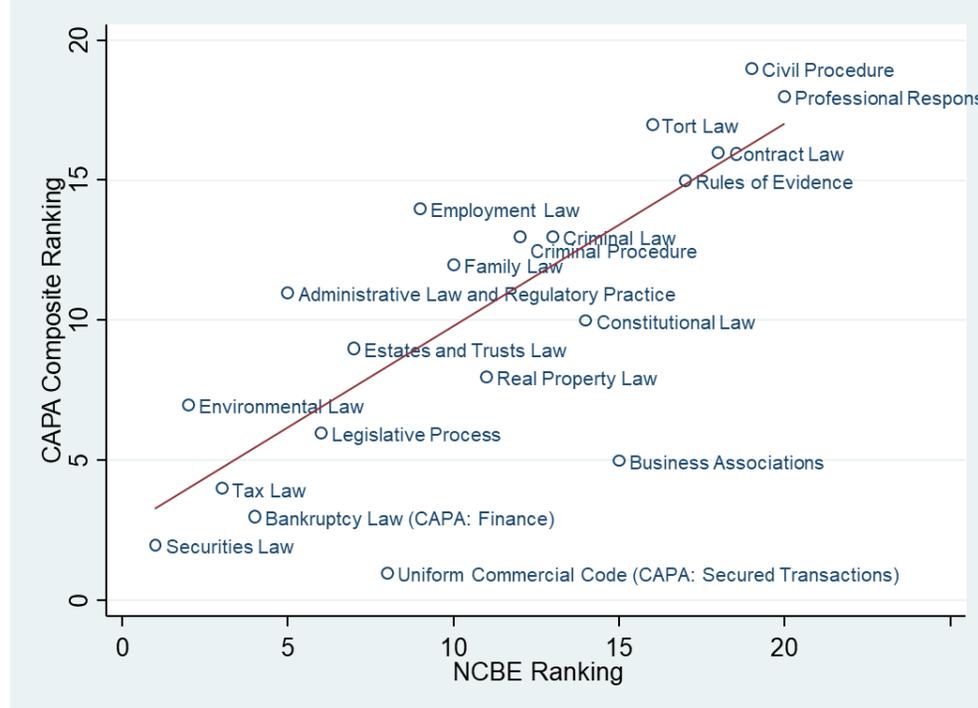
¹¹ The missing NCBE items are Legal Research Methodology (ranked 5), Statutes of Limitations (6), Local Court Rules (7), Statutory Interpretation Principles (8), Sources of Law (9).

Table 8. Comparison between CAPA and NCBE Practice Analysis Study on Legal Topics

NCBE Knowledge Areas	NCBE Ranking (Mean Importance)	CAPA Composite Ranking
Civil Procedure	2	1
Professional Responsibility	1	2
Tort Law	10	3
Contract Law	3	4
Rules of Evidence	4	5
Employment Law	22	6
Criminal Law	15	7
Criminal Procedure	16	7
Family Law	21	8
Administrative Law and Regulatory Practice	32	9
Constitutional Law	13	10
Estates and Trusts Law	25	11
Real Property Law	17	12
Environmental Law	54	13
Legislative Process	27	14
Business Organization Law (CAPA: Business Associations)	11	15
Tax Law	47	16
Bankruptcy Law (CAPA: Finance)	46	17
Securities Law	56	18
Uniform Commercial Code (CAPA: Secured Transactions)	24	19

To further assess the degree of correlation between the CAPA and NCBE rankings, Figure 3 transformed the rank orders slightly, first by eliminating the gaps in NCBE’s ranking, and then reversing the numerical orders for both so that items with higher ranking is located at a higher position (Civil Procedure’s number 1 ranking transformed to 19, for example). With the exception of a few items misaligned, the results from the two surveys reveal a remarkable correlation (correlation coefficient of 0.77).

Figure 3. Highly Correlated Rankings in Legal Topics from CAPA and NCBE Results



Conclusion

The detailed, robust data from two surveys conducted under the guidance of the CAPA Working Group provides an empirical foundation for reaching consensus on the legal topics that should be covered on the CBX as well as the competencies expected of entry-level attorneys. This work will be invaluable for ensuring that the Exam is adapted to reflect the current practice of law in California.

Appendix A. CAPA Working Group Members

Alan Steinbrecher	State Bar of California Board of Trustees	Chair
Robert Brody	Committee of Bar Examiners	Member
Shanae Buffington	Employment Development Department	Attorney
Sarah Cassady	Young, Minney & Corr, LLP	Associate Attorney
Margie Estrada	California Senate Judiciary Committee	Chief Counsel
Stephen Ferruolo	University of San Diego School of Law	Dean
Jackie Gardina	Santa Barbara and Ventura College of Law	Dean
David George	Geosyntec Consultants, Inc.	Corporate Counsel
Sylvia Kennedy	Senior Adults Legal Assistance (SALA)	Senior Staff Attorney
Judge Cindee Mayfield	Mendocino Superior Court	Judge
Dr. Justin McCrary	Columbia Law School	Professor
Gregory Murphy	National Conference of Bar Examiners	Trustee (former)
John Palmer	Orrick	Partner
Judge Glen M. Reiser	Ventura Superior Court	Judge (Ret.)
Emily Scivoletto	UC Davis School of Law	Senior Assistant Dean
Dr. Aaron Taylor	AccessLex Institute	Senior Vice President

Appendix B. Practice Analysis Study Panel Members

Robert Bailey	Keri Klein	Laura Palazzolo
Robert Barrett	Lydia Liberio	Yan Shrayberman
Shanae Buffington	Kwixuan Maloof	Sheniece Smith
Stephen Ferruolo	Sean McCoy	Sabrina Thomas
Dustin Johnson	Laura Nelson	Marcus Wiggins

Appendix C. List of Competencies, Tasks, Legal Topics, and Subtopics

The State Bar has designed a two-way framework to link performance in practice to content. **Areas of Responsibilities** are the major responsibilities or duties that characterize the practice of law. For each of the four areas of responsibility there are three competencies. **Legal Topics** are major topics in law. They are augmented with more specific topics, or subtopics, that offer detail about the legal topic.

Areas of Responsibility	Legal Topics
Establishing and Maintaining Relationships <ul style="list-style-type: none"> Establishing the client relationship Maintaining the client relationship Communicating with others 	Administrative Law and Procedure Business Associations Civil Procedure Constitutional Law Contracts
Practice Management and Administration <ul style="list-style-type: none"> Practice management Case or matter management Supervision and collaboration 	Criminal Law and Procedure Environment and Land Use Estate Planning, Trusts, and Probate Evidence Family Law
Factual and Legal Analysis <ul style="list-style-type: none"> Issue spotting and fact gathering Research and investigation Drafting and writing 	Finance Labor and Employment Legislation Professional Responsibility Real Property
Resolutions <ul style="list-style-type: none"> Counsel/advice (via conversation, in-person or by telephone) Negotiation and closing Alternative Dispute Resolution Litigation 	Secured Transactions Securities Tax Torts

On the following pages are the tasks that attorneys perform when they work in one of the competency domains. Overall, there are 117 tasks that apply to attorneys in any practice area.

Establishing and Maintaining Relationships

Establishing the client relationship

1. Identify the client(s)
2. Assess potential conflicts of interest
3. Manage conflicts throughout representation
4. Determine the client's goals and expectations
5. Evaluate competence to represent the client's interests
6. Manage referrals to and from other attorneys
7. Define the scope of the attorney-client relationship
8. Explain the client's obligations and responsibilities
9. Manage third-party involvement in representation of the client(s)
10. Document the engagement (e.g., engagement letter, arbitration agreement, fee agreement, conflict waiver)
11. Document the decision to decline representation

Maintaining the client relationship

1. Update the client(s) throughout the matter
2. Respond to client inquiries
3. Resolve disputes with clients
4. Document termination of the representation

Communicating with others

1. Determine disclosure or notice obligations
2. Determine confidentiality obligations
3. Communicate with opposing counsel or parties
4. Communicate with other interested persons (e.g., media, regulatory bodies, insurers)
5. Manage communications with other interested persons (e.g., media, regulatory bodies, insurers)
6. Communicate with witnesses, consultants, or experts
7. Manage communications with witnesses, consultants, or experts

Practice Management and Administration

Practice management

1. Determine necessary staffing
2. Manage staff members
3. Comply with State Bar licensing and reporting requirements
4. Create the practice's budget
5. Manage the practice's finances
6. Manage client trust accounts
7. Manage appropriate insurance coverage
8. Manage IT requirements and resources
9. Manage service vendors
10. Market the practice in compliance with requirements relating to attorney advertising
11. Manage files and records
12. Maintain calendar

Case or matter management

1. Establish the budget for the case or matter
2. Calendar deadlines
3. Record time spent on the case or matter
4. Manage client billing

Supervision and collaboration

1. Seek advice from senior attorney(s)
2. Delegate tasks
3. Oversee delegated tasks
4. Define expectations and scope of responsibilities between co-counsels

Factual and Legal Analysis**Issue-spotting and fact-gathering**

1. Interview the client
2. Obtain documents and evidence from the client
3. Obtain public information about the client and the matter
4. Review the documents collected
5. Identify legal and factual issues
6. Identify other persons with relevant information or interests
7. Prepare a summary of the interview and initial assessment of the client

Research and investigation

1. Determine governing laws
2. Research laws and precedents
3. Consult with attorneys who have more experience or expertise
4. Evaluate exemplars, forms, and models
5. Search for relevant records and documents
6. Review relevant records and documents
7. Determine the need for non-attorney consultants or experts
8. Confer with non-attorney consultants or experts
9. Interview persons with possibly relevant information or interests
10. Prepare internal working analysis of the case or matter

Drafting and writing

1. Draft correspondence
2. Use exemplars, forms, and models
3. Draft pleadings, motions, statements, or briefs
4. Draft formal legal instruments (e.g., liens, claims, title documents, estate or incapacity plans, settlement agreements)
5. Draft discovery requests or responses
6. Draft legal memoranda
7. Draft contracts and other transactional documents
8. Redline transactional documents
9. Edit drafts or documents

Resolutions

Counsel/Advice

1. Advise the client regarding the benefits, risks, and consequences of an issue and/or course of action
2. Advise the client about behavioral expectations
3. Advise the client on specific legal questions and rules
4. Document communications and advice given to the client
5. Conduct necessary follow up

Negotiation and Closing

1. Explain the terms, conditions, and status of negotiations
2. Coordinate the roles and authority of participants in negotiations
3. Participate in negotiations
4. Represent the client in mediation of transactional disputes
5. Coordinate closing of a transaction

Alternative Dispute Resolution

1. Evaluate options for alternative dispute resolution
2. Evaluate potential neutrals
3. Communicate with neutrals
4. Prepare for alternative dispute resolution proceedings
5. Represent the client in alternative dispute resolution proceedings

Litigation

1. Analyze jurisdictional issues
2. Analyze proper venue and statute of limitations
3. Research local rules
4. Identify parties' legal names and capacities
5. Comply with statutory notice and service requirements
6. Analyze discovery needs
7. Develop discovery plan
8. Implement discovery plan
9. Analyze e-discovery requirements and obligations
10. Instruct the client regarding the preservation of evidence
11. Instruct the client regarding the production of evidence pursuant to discovery requests
12. Review evidence for production
13. Prepare for depositions
14. Prepare witnesses to testify
15. Attend depositions
16. Review preliminary hearing record
17. Review deposition transcripts
18. Summarize deposition transcripts
19. Determine motion strategy and file relevant motions
20. Appear at hearings
21. Prepare for trial (e.g., subpoenas, exhibits, motions in limine, jury instructions)
22. Appear at trial
23. Prepare post-trial motions

24. Appear at post-trial hearings
25. Evaluate potential appeal
26. File notice of appeal

The legal topics address topics in law about which attorneys are expected to be knowledgeable, depending on the area of practice in which they are engaged. There are 19 legal topics, each includes a number of subtopics as listed on the following pages. Within each subtopics are listed several major sources related to the subtopic.

Administrative Law and Procedure

Rule Making

1. Administrative Procedure Act
2. California Code of Regulations
3. Code of Federal Regulation
4. Case law

Interpretation

1. Case law
2. Agency precedential decisions and opinions

Regulatory Enforcement

1. Administrative Procedure Act
2. California Code of Regulations
3. Code of Federal Regulation
4. Case law

Agency Claims and Hearings

1. Administrative Procedure Act
2. California Code of Regulations
3. Code of Federal Regulation
4. Case law
5. Agency precedential decisions and opinions

Judicial Review and Appeals

1. Administrative Procedure Act
2. California Code of Regulations
3. Code of Federal Regulation
4. Case law

Business Associations

Types of Business Entities

1. California Corporations Code

Formation and Governing Documents

1. California Corporations Code

Stock, Membership Interests

1. California Corporations Code

Governance

1. California Corporations Code
2. Case law

Fiduciary Duties

1. California Corporations Code
2. Case law

Corporate Powers, Ultra Vires

1. California Corporations Code
2. Case law

Limitations of Liability/Piercing Corporate Veil

1. California Corporations Code
2. Case law

Merger and Acquisition, Dissolution, Winding Up

1. California Corporations Code
2. Case law

Civil Procedure

Jurisdiction

1. California Code of Civil Procedure
2. Rules of Court
3. Federal Rules of Civil Procedure
4. Case law

Pleadings

1. California Code of Civil Procedure
2. Rules of Court
3. Federal Rules of Civil Procedure
4. Case law

Motions

1. California Code of Civil Procedure
2. Rules of Court
3. Federal Rules of Civil Procedure
4. Case law

Discovery

1. California Code of Civil Procedure
2. Rules of Court
3. Federal Rules of Civil Procedure
4. Case law

Trials

1. California Code of Civil Procedure
2. Rules of Court
3. Federal Rules of Civil Procedure
4. Case law

Appeals

1. California Code of Civil Procedure
2. Rules of Court
3. Federal Rules of Civil Procedure
4. Case law

Enforcement of Judgments

1. California Code of Civil Procedure
2. Enforcement of judgments law
3. Case law

Constitutional Law

First Amendment

1. U.S. Constitution
2. Case law

Fourth Amendment

1. U.S. Constitution
2. Case law

Fifth Amendment

1. U.S. Constitution
2. Case law

Sixth Amendment

1. U.S. Constitution
2. Case law

Eighth Amendment

1. U.S. Constitution
2. Case law

Commerce Clause

1. U.S. Constitution
2. Case law

Separation of Powers

1. U.S. Constitution
2. Case law

Federalism

1. U.S. Constitution
2. Case law

Due Process and Equal Protection

1. U.S. Constitution
2. Case law

Sovereign Immunity

1. U.S. Constitution
2. Case law

Contracts

Offer and Acceptance

1. Uniform Commercial Code
2. California Civil Code
3. Common law

Modification, Amendment, Novation

1. Uniform Commercial Code
2. California Civil Code
3. Common law

Performance

1. Uniform Commercial Code
2. California Civil Code
3. Common law

Interpretation

1. Uniform Commercial Code
2. California Civil Code
3. Common law

Consideration

1. Uniform Commercial Code
2. California Civil Code
3. Common law

Enforceability

1. Uniform Commercial Code
2. California Civil Code
3. Common law

Breaches and Remedies

1. Uniform Commercial Code
2. California Civil Code
3. Common law

Effectiveness, Term, and Termination

1. Uniform Commercial Code
2. California Civil Code
3. Common law

Criminal Law and Procedure

Crimes (e.g., person, property)

1. California Penal Code

Criminal Procedure

1. Federal constitution
2. State constitution
3. Statutory law
4. Case law

Sentencing

1. California Penal Code
2. Federal constitution
3. State constitution
4. Statutory law
5. Case law

Pre-Trial Writs

1. Federal constitution
2. State constitution
3. Statutory law
4. Case law

Plea Negotiation

1. Federal constitution
2. State constitution
3. Statutory law
4. Case law

Trial

1. Federal constitution
2. State constitution
3. Statutory law
4. Case law

Appeal and Post-Trial Writs

1. Case law

Environment and Land Use

Due Diligence for Acquisition

Environmental Impact Statements

1. California Environmental Quality Act & Guidelines
2. National Environmental Policy Act & Guidelines
3. Case law implementing relevant statutes

Eminent Domain

1. Code of Civil Procedure
2. Constitutional Takings Doctrine (5th Amendment)

Remediation

Regulatory Compliance

Hazardous Waste and Toxic Substances Control and Remediation

1. Resource Conservation and Recovery Act

2. Comprehensive Environmental Response, Compensation, and Liability Act
3. Porter-Cologne Act
4. Toxic Substances Control Act
5. California Hazardous Waste Control Act
6. Proposition 65
7. Oil Pollution Act

Entitlements

1. Public records
2. Planning and Zoning Laws (local and state)
3. California Coastal Act
4. California Environmental Quality Act
5. Clean Water Act section 404
6. California Public Trust Doctrine

Water Rights

1. Clean Water Act
2. Clean Air Act

Estate Planning, Trusts, and Probate

Estate Planning

1. California Probate Code
2. Internal Revenue Code
3. Revenue and Taxation Code
4. California Family Law Code

Guardianship and Probate Conservatorship

1. California Probate Code

Administration of Trusts

1. California Probate Code
2. Internal Revenue Code
3. Revenue and Taxation Code
4. California Family Law Code

Administration of Decedents' Estates

1. California Probate Code
2. Internal Revenue Code
3. Revenue and Taxation Code
4. California Family Law Code

Nonprobate Transfers

1. California Probate Code
2. Internal Revenue Code
3. Revenue and Taxation Code
4. California Family Law Code

Evidence

Admissibility

1. Federal rules of evidence
2. Evidence Code
3. Case law

Privileges

1. Federal rules of evidence
2. Evidence Code
3. Case law

Family Law

Division of Property

1. California Family Law Code
2. Common law

Support

1. California Family Law Code
2. Internal Revenue Code

Custody

1. California Family Law Code
2. California Penal Code

Documents

1. Federal rules of evidence
2. Evidence Code
3. Case law

Burdens and Presumptions

1. Federal rules of evidence
2. Evidence Code
3. Case law

3. Welfare and Institutions Code
4. Probate Code, Division 4

Adoption

1. California Family Law Code

Domestic Partnership

1. California Family Law Code

Marriage and Dissolution

1. California Family Law Code

Finance

Principles of Finance and Valuation

1. GAAP
2. Uniform Commercial Code
3. California Financial Code
4. California Commercial Code
5. California Corporations Code
6. State statutes and regulations

Capital Structure

1. Internal Revenue Code
2. Securities law
3. Securities Act of 1933
4. Securities and Exchange Act of 1934
5. Investment Advisors Act of 1940

Financial Instruments

1. Uniform Commercial Code
2. California Financial Code
3. California Commercial Code

Fixed Income Markets

1. Securities and Exchange Act
2. Securities Act of 1933
3. Internal Revenue Code

Bankruptcy

1. Bankruptcy Code

Securities and Securities Markets

1. Securities and Exchange Act of 1934
2. Securities Act of 1933

Financial Reporting and Disclosure

1. GAAP
2. Federal statutes and regulations
3. State statutes and regulations
4. Securities law
5. Securities Act of 1933
6. Securities and Exchange Act of 1934
7. Governmental Accounting Standards Board
8. Financial Accounting Standards Board
9. Public Company Accounting Oversight Board guidance documents
10. International Accounting Standards Board

Fraud, Corruption, and Regulatory Risk

1. GAAP
2. Federal statutes and regulations
3. State statutes and regulations
4. Case law

Labor and Employment

Classification of Employee

1. Statutes
2. Case law
3. Regulations

Wage and Hour

1. Statutes
2. Case law
3. Regulations

Discrimination and Harassment

1. Statutes
2. Case law
3. Regulations

Hiring and Termination

1. Statutes
2. Case law
3. Regulations

Human Resource Policies

1. Statutes
2. Case law
3. Regulations

Labor Organizations

1. Statutes
2. Case law
3. Regulations

Legislation and Government

Legislative Process

1. California Constitution
2. Lobbying Disclosure Manual

Legislative History

1. Westlaw
2. Lexis

Lobbying

1. Lobbying Disclosure Information Manual
2. Political Reform Act
3. Fair Political Practices Commission Regulations

4. Fair Political Practices Commission Advice Letters and Commission Opinions
5. Conflict of Interest Code Exemptions

Constitutional Bases

1. California Constitution
2. U.S. Constitution

Government

Municipal Law

Professional Responsibility

Competence

1. American Bar Association Model Rules
2. California Rules of Professional Conduct
3. Business and Professions Code
4. Common law tort of negligence
5. Statutory standards of care

Confidentiality

1. American Bar Association Model Rules
2. California Rules of Professional Conduct
3. Business and Professions Code
4. Evidence Code
5. Code of Civil Procedure

Conflicts of Interest

1. American Bar Association Model Rules

2. California Rules of Professional Conduct
3. Business and Professions Code

Fees, Billing, and Trust Accounting

1. American Bar Association Model Rules
2. California Rules of Professional Conduct
3. Business and Professions Code

Advocacy

1. American Bar Association Model Rules
2. California Rules of Professional Conduct
3. Business and Professions Code
4. Rules of Court

Real Property

Estates, Easements, and Future Interests

1. California Civil Code
2. California Probate Code
3. Common law
4. Case law

Landlord Tenant

1. California Civil Code
2. Common law
3. Statutory law
4. Rent stabilization orders

Conveyances

1. California Civil Code
2. Common law

Title

1. Common law
2. Statutory law

Secured Transactions

Real Property Liens

1. California Civil Code

Personal Property Liens

1. Uniform Commercial Code

Enforcement and Foreclosure

1. California Civil Code
2. Uniform Commercial Code

Securities

Definition of Security

1. Case law

Public Offerings

1. Federal statutes and regulations

Exemptions and Private Offerings

1. Federal statutes and regulations
2. Blue sky laws

Securities Markets and Professionals

1. Federal statutes and regulations

Insider Trading

1. Federal statutes and regulations
2. Case law

Reporting and Regulatory Compliance

1. Federal statutes and regulations
2. Case law

Tax

Income Taxation

1. Internal Revenue Code
2. California Revenue and Taxation Code

Corporate and Partnership Taxation

1. Internal Revenue Code
2. California Revenue and Taxation Code

Property Taxation

1. Internal Revenue Code
2. California Revenue and Taxation Code

Estate and Gift Taxation

1. Internal Revenue Code

Torts

Negligence

1. California Civil Code
2. Common law

Intentional Torts

1. California Civil Code
2. Common law

Strict Liability

1. California Civil Code
2. Common law

Products Liability

1. California Civil Code
2. Common law

Remedies

1. California Civil Code
2. Common law

Attachment B



**State Bar of California
Differential Item Functioning Analysis Report**

December 2019

SCANTRON®

TABLE OF CONTENTS

INTRODUCTION.....	1
METHODS	2
Sample	2
DIF Identification Methods.....	3
DIF Flagging Criteria	4
RESULTS	5
Gender	5
Race/Ethnicity	7
Law School Type.....	9
Gender, Race/Ethnicity, and Law School Type.....	11
CONCLUSION	13
Gender	13
Race/Ethnicity	13
Law School Type.....	14
Gender, Race/Ethnicity, and Law School Type.....	14
Recommendations	15
REFERENCES.....	17

TABLES

Table 1. Sample Demographic Information
Table 2. Essay and Performance Test Item Score Strata
Table 3. Candidate MBE Score Strata
Table 4. Example MH Table
Table 5. DIF Flagging Criteria
Table 6. Example Results Table
Table 7. Gender DIF Statistics
Table 8. Gender and Item Type DIF Statistics
Table 9. Gender and Subject DIF Statistics
Table 10. Race/Ethnicity DIF Statistics
Table 11. Race/Ethnicity and Subject DIF Statistics
Table 12. Law School Type DIF Statistics
Table 13. Law School and Administration Window DIF Statistics
Table 14. Gender, Race/Ethnicity, and Law School Type

APPENDICES

APPENDIX A: GENDER ITEM-LEVEL DIF RESULTS
APPENDIX B: RACE/ETHNICITY ITEM-LEVEL DIF RESULTS
APPENDIX C: LAW SCHOOL TYPE ITEM-LEVEL DIF RESULTS
APPENDIX D: RACE/ETHNICITY BY SUBJECT RESULTS

INTRODUCTION

Created by the Legislature in 1927, the State Bar of California is an arm of the California Supreme Court, protecting the public by licensing and regulating attorneys. It serves the people of California through careful oversight of the legal profession.

An attorney must pass several examinations to be licensed to practice law in California by the State Bar. The California Bar Examination (CBX), administered twice each year in February and July, consists of the following parts:

- Essay questions,
- Performance test, and
- Multiple-choice questions via the Multistate Bar Examination (MBE).

Since 2009, the examination has covered the following subjects: Business Associations, Civil Procedure, Community Property, Constitutional Law, Contracts, Criminal Law and Procedure, Evidence, Professional Responsibility, Real Property, Remedies, Torts, Trusts, and Wills and Succession.

The State Bar contracted with Scantron Corporation, a full-service certification and licensure company, to conduct a differential item functioning (DIF) analysis for essay and performance test items administered as part of the CBX between 2009 and 2019. The purpose of a DIF analysis is to assess whether different groups with similar overall scores perform differently on test items.

This report summarizes the findings of a DIF study which considered the item performance based on the following primary variables: gender, race/ethnicity, and law school type. Secondary variables, such as item type, subject matter, and administration window were also considered in the analysis.

METHODS

Sample

First-time candidates who were administered the CBX between July 2009 and February 2019 were included in this analysis. A total of 116 essay and 36 performance test items were included in this study. Both item types were scored on a scale from 40 to 100 with increments of five. All candidates in the analysis had an MBE scaled score, which was used as an indicator of overall performance for DIF purposes.

The demographic variables included in the DIF analysis were gender (Male, Female), race/ethnicity (Asian, Black/African American, Hispanic, Caucasian, Other), and law school type (California ABA, California Accredited, California Registered, Out-of-state ABA, Attorney, Foreign JD). The number of first-time candidates included from each administration window is included in Table 1 below.

Table 1. Sample Demographic Information

Item	Gender		Race/Ethnicity					Law School Type					
	Male	Female	Asian	Black/AA	Hispanic	Caucasian	Other	CA ABA	CA Accredited	CA Registered	Out-of-state ABA	Attorney	Foreign JD
G0907	2472	2153	914	121	315	2974	58	3026	137	53	1098	241	76
G1002	432	423	170	23	59	551	11	300	41	46	138	289	41
G1007	3073	2991	1291	208	548	3559	91	3763	279	110	1351	355	193
G1102	499	396	166	28	56	574	14	318	37	49	147	294	50
G1107	2432	2027	863	130	353	2826	68	3053	101	34	955	243	73
G1202	463	444	171	36	68	569	12	309	42	43	166	286	62
G1207	2552	2197	883	139	374	3070	42	3237	109	37	1010	279	81
G1302	476	390	173	42	55	525	14	292	32	45	145	284	68
G1307	2576	2256	982	125	402	2991	72	3264	137	38	996	279	119
G1402	554	432	189	37	70	604	10	348	58	41	148	323	68
G1407	2249	1963	838	152	377	2548	64	2734	133	46	888	298	114
G1502	459	441	179	51	81	487	18	254	54	47	133	316	96
G1507	1934	1919	828	143	397	2334	56	2511	89	35	859	265	130
G1506	376	401	172	31	68	487	16	217	37	35	125	292	83
G1607	1585	1629	664	104	348	1979	51	2132	73	38	687	195	127
G1702	282	311	137	35	71	337	11	184	43	26	83	187	80
G1707	1820	2016	871	170	433	2208	61	2281	129	38	819	330	273
G1802	386	423	188	47	88	464	20	198	57	43	111	269	158
G1807	2338	2640	1215	284	762	2563	87	3083	228	68	916	299	462
G1902	439	485	256	52	123	477	9	244	76	59	123	254	193

Candidate item scores and MBE results were stratified in order to have enough candidates across each distribution for the DIF analyses. Item scores were grouped into a single stratum at the extreme ends of the 40 to 100 scale, with scores of 50 and lower grouped into the lowest level and scores of 85 and higher grouped into the highest level. The stratification rules are displayed below in Table 2.

Table 2. Essay and Performance Test Item Score Strata

Rescaled Score	Item Score
1	50 or lower
2	55
3	60
4	65
5	70
6	75
7	80
8	85 or higher

Similarly, MBE scores lower than 1200 were grouped into the lowest stratum and scores of 1700 and higher were grouped into the highest stratum. Other strata were created using intervals of 100. The strata rules are displayed below in Table 3.

Table 3. Candidate MBE Score Strata

Stratum for Study	MBE Score
1	Lower than 1200
2	1200-1299
3	1300-1399
4	1400-1499
5	1500-1599
6	1600-1699
7	1700 or higher

DIF Identification Methods

Two methods were used to assess DIF in this study:

1. The generalized Mantel-Haenszel (MH) method, and
2. The Liu-Agresti estimator of the cumulative common odds ratio (LA-LOR).

The MH method is an DIF analysis tool that detects item response differences across subgroups (e.g., gender, race/ethnicity), while controlling for overall performance. The MH statistic is distributed as a chi-square statistic with one degree of freedom. Under the MH method, data are organized into a $2 \times T \times K$ contingency table, where T represents the number of response categories and K represents the number of levels of a stratification variable for each of the two groups. For an item with four response categories, the data for each stratum would be represented in a table like the one below in Table 4, where n_{RTK} and n_{FTK} represent the reference and focal groups for each stratum ($K = 1, 2, 3, 4$).

Table 4. Example MH Table

	Response 1	Response 2	Response 3	Response 4	Total
Reference group	n_{R1K}	n_{R2K}	n_{R3K}	n_{R4K}	
Focal group	n_{F1K}	n_{F2K}	n_{F3K}	n_{F4K}	
Total					N_K

The LA-LOR statistic is analogous to the Mantel-Haenszel common log odds ratio (MH-LOR) used for dichotomous items. Like the MH-LOR, the LA-LOR provides an overall assessment of DIF for an item, with positive values indicating DIF in favor of the reference group and negative values indicating DIF in favor of the focal group.

The MH and LA-LOR statistics were calculated using SAS version 9.4 (SAS Institute, 2014) and DIFAS version 5.0 (Penfield, 2013).

DIF Flagging Criteria

In this study both the MH and LA-LOR statistics were used to identify DIF. Items were flagged under the MH method for statistically significant DIF based on a critical value of 3.84, which corresponds to a probability level of 0.05. Although statistical significance is helpful, an effect size for observed DIF was also desired.

Therefore, the LA-LOR statistic was used in this study to identify three levels of DIF. Observed DIF was categorized as either negligible (at or below 0.43), moderate (between 0.43 and 0.64), or large (at or above 0.64) based on the LA-LOR absolute value. The method is similar to the widely used classification scheme developed by Educational Testing Service (Zieky, 1993) for use with dichotomous items and implemented by Penfield and Algina (2006) in a DIF study similar to this one, which included a combination of dichotomous and polytomous items.

The focus of this study was on items that both have statistically significant DIF, as identified by the MH method, and were identified as having moderate or large DIF using the LA-LOR statistic. The flagging criteria are outlined in Table 5.

Table 5. DIF Flagging Criteria

DIF Statistics	DIF Flagging Criteria
MH Statistic	Critical value of 3.84
LA-LOR Absolute Value	Moderate DIF (Between 0.43-0.64) or Large DIF (Greater than or equal to 0.64)

RESULTS

The results section focuses on trends found with respect to three specific DIF variables of interest: gender, race/ethnicity, and law school type. Secondary variables, such as item type, subject matter, and administration window are also presented where appropriate.

DIF results are presented in aggregate in tables which display the MH and LA-LOR DIF identification rules. An example is provided below in Table 6. The numbers of flagged items are bolded and are located in the lower left and lower right areas of the table.

Table 6. Example Results Table

MH Statistic	Favors Focal Group			Favors Reference Group		
	Large	Moderate	Negligible	Negligible	Moderate	Large
Nonsig.	# of items	# of items	# of items	# of items	# of items	# of items
Significant	# of items	# of items	# of items	# of items	# of items	# of items

In addition to the aggregate trends discussed in body of this report, item-level DIF statistics are included in Appendices A (gender), B (race/ethnicity), and C (law school type).

Gender

Males were the reference group and females were the focal group for the gender analysis. Negative LA-LOR statistics were observed for 150 of 152 items, indicating that females performed better than males on 99% of items after controlling for overall performance on the MBE. Gender DIF analysis results are provided below in Table 7.

Table 7. Gender DIF Statistics

MH Statistic	LA-LOR Statistic						
	Large	Favors Female			Favors Male		
		Moderate	Negligible	Negligible	Moderate	Large	
Nonsig.	0	0	37	2	0	0	
Significant	1	29	83	0	0	0	

Although a total of 113 items were flagged by the MH statistics for a significant level of DIF, the level of DIF was categorized as negligible for 83 of those items. DIF was categorized as moderate or large for only 30 items, or 20%, and those items are bolded along the bottom left side of Table 7.

As a follow-up analysis, the frequency of items flagged for gender DIF for each item type was examined. The results of the analysis are provided below in Table 8.

Table 8. Gender and Item Type DIF Statistics

Item Type	MH Statistic	LA-LOR Statistic						
		Large	Favors Female			Favors Male		
			Moderate	Negligible	Negligible	Moderate	Large	
Essay	Nonsig.	0	0	34	2	0	0	
	Significant	1	20	59	0	0	0	
Performance Test	Nonsig.	0	0	3	0	0	0	
	Significant	0	9	24	0	0	0	

The proportion of items flagged for DIF was similar for essay (21 items out of 116, or 18%) and performance test (9 items out of 36, or 25%) items.

The frequency of items flagged for gender DIF for each subject was also examined. Table 9 provides a breakdown of the analysis.

Table 9. Gender and Subject DIF Statistics

Subject	MH Statistic	Large	LA-LOR Statistic				Large
			Favors Female		Favors Male		
			Moderate	Negligible	Negligible	Moderate	
Business	Nonsig.	0	0	3	0	0	0
Associations	Significant	0	1	7	0	0	0
Civil	Nonsig.	0	0	1	1	0	0
Procedure	Significant	0	3	9	0	0	0
Community	Nonsig.	0	0	1	0	0	0
Property	Significant	0	5	5	0	0	0
Constitutional	Nonsig.	0	0	5	1	0	0
Law	Significant	0	0	4	0	0	0
Contracts	Nonsig.	0	0	5	0	0	0
	Significant	0	4	16	0	0	0
Criminal Law	Nonsig.	0	0	5	0	0	0
& Procedure	Significant	0	0	7	0	0	0
Evidence	Nonsig.	0	0	1	0	0	0
	Significant	0	4	6	0	0	0
Immigration	Nonsig.	0	0	0	0	0	0
Law*	Significant	0	0	1	0	0	0
Professional	Nonsig.	0	0	4	0	0	0
Responsibility	Significant	0	5	7	0	0	0
Real Property	Nonsig.	0	0	3	0	0	0
	Significant	0	0	8	0	0	0
Regulatory	Nonsig.	0	0	1	0	0	0
Law*	Significant	0	1	0	0	0	0
Remedies	Nonsig.	0	0	3	0	0	0
	Significant	0	0	3	0	0	0
Torts	Nonsig.	0	0	4	0	0	0
	Significant	0	1	7	0	0	0
Trusts	Nonsig.	0	0	0	0	0	0
	Significant	0	2	1	0	0	0
Wills &	Nonsig.	0	0	1	0	0	0
Succession	Significant	1	3	2	0	0	0

*Immigration Law and Regulatory Law subject areas were tested via performance test items only and therefore are not listed as subjects tested on the CBX exam.

The legal subjects flagged for gender DIF the most based on proportion of items were Trusts (2 of 3, or 67%), Wills and Succession (4 of 7 items, or 57%), Regulatory Law (1 of 2, or 50%), and Community Property (5 of 11 items, or 45%). No items were flagged for gender DIF from the following subjects: Constitutional Law, Criminal Law and Procedure, Immigration Law, Real Property, and Remedies.

Race/Ethnicity

A number of DIF analyses were conducted with respect to the race/ethnicity variables. In all cases, Caucasian was designated as the reference variable and the focal groups for each analysis were Black/African American, Asian, Hispanic, or Other.

In some administration windows, the difference between the number of candidates in the reference and focal groups was quite large (see Table 1). For instance, the number of Black/African American and Other candidates was often much smaller compared to the number of candidates in the Caucasian reference group. Large differences in sample size may have impacted the results, potentially resulting in flagging items for race/ethnicity DIF at a higher or lower rate than if the groups had been of similar size.

The most concerning race/ethnicity categories with respect to small sample sizes were Black/African American, with as few as 23 candidates in some administration windows, and Other, with as few as nine candidates in some administration windows. The statistical instability of using sample sizes this small was cause for concern.

Therefore, a variable more comparable in sample size to the Caucasian reference group was created by combining the Black/African American, Asian, Hispanic, and Other categories into a single variable, labeled as Non-Caucasian in Table 10, Table 11, and Appendix B. The purpose of this variable was to include valuable information from the Black/African American and Other categories while avoiding concerns related to small sample sizes.

Since the Non-Caucasian variable produced sample sizes for the focal group that were more similar in size compared to the Caucasian reference group, it was used as the primary race/ethnicity variable for follow-up analyses. Follow-up analyses for each specific race/ethnicity variable were also conducted and are included in Appendix D, although the Black/African American and Other analyses should be interpreted with caution due to sample size concerns.

All race/ethnicity DIF analyses are included in Table 10.

Table 10. Race/Ethnicity DIF Statistics

Focal Group	MH Statistic	LA-LOR Statistic					
		Favors Focal Group			Favors Caucasian		
		Large	Moderate	Negligible	Negligible	Moderate	Large
Black/African American	Nonsig.	0	1	11	71	22	1
	Significant	0	0	0	21	15	10
Asian*	Nonsig.	0	0	50	62	0	0
	Significant	0	0	11	18	4	6
Hispanic*	Nonsig.	0	0	55	73	0	0
	Significant	0	0	1	14	4	3
Other	Nonsig.	2	8	43	70	8	14
	Significant	0	0	0	0	5	2
Non- Caucasian	Nonsig.	0	0	36	79	0	0
	Significant	0	0	2	27	8	0

*Observed items with MH statistics and LA-LOR statistics of zero, therefore not all 152 items appear in the table

Even with the large sample size discrepancies, very few items were flagged for race/ethnicity DIF. The largest number of items was flagged in the Black/African American DIF analysis (25 items or 16%), followed by the Asian (10 items or 7%), Non-Caucasian (8 items or 5%), Hispanic (7 items or 5%), and Other (7 items or 5%) analyses which are highlighted in Table 10. Although few items were flagged for race/ethnicity DIF, all flagged items indicated that Caucasians performed better than each focal group after controlling for overall performance on the MBE.

The frequency of items flagged for race/ethnicity DIF for each subject was also examined. Table 11 provides a breakdown of the analysis for the Non-Caucasian focal group. The analyses for the Black/African American, Asian, Hispanic, and Other focal groups are included in Appendix D.

Table 11. Race/Ethnicity and Subject DIF Statistics

Subject	MH Statistic	LA-LOR Statistic					
		Favors Non-Caucasian			Favors Caucasian		
		Large	Moderate	Negligible	Negligible	Moderate	Large
Business	Nonsig.	0	0	6	4	0	0
Associations	Significant	0	0	0	1	0	0
Civil	Nonsig.	0	0	9	3	0	0
Procedure	Significant	0	0	0	2	0	0
Community	Nonsig.	0	0	2	9	0	0
Property	Significant	0	0	0	0	0	0
Constitutional	Nonsig.	0	0	2	7	0	0
Law	Significant	0	0	0	1	0	0
Contracts	Nonsig.	0	0	4	14	0	0
	Significant	0	0	1	5	1	0
Criminal Law	Nonsig.	0	0	2	3	0	0
& Procedure	Significant	0	0	0	6	1	0
Evidence	Nonsig.	0	0	1	7	0	0
	Significant	0	0	0	2	1	0
Immigration	Nonsig.	0	0	0	1	0	0
Law*	Significant	0	0	0	0	0	0
Professional	Nonsig.	0	0	2	8	0	0
Responsibility	Significant	0	0	0	5	1	0
Real Property	Nonsig.	0	0	2	6	0	0
	Significant	0	0	0	1	2	0
Regulatory	Nonsig.	0	0	1	1	0	0
Law*	Significant	0	0	0	0	0	0
Remedies	Nonsig.	0	0	0	5	0	0
	Significant	0	0	0	1	0	0
Torts	Nonsig.	0	0	3	4	0	0
	Significant	0	0	1	2	2	0
Trusts	Nonsig.	0	0	1	2	0	0
	Significant	0	0	0	0	0	0
Wills &	Nonsig.	0	0	1	5	0	0
Succession	Significant	0	0	0	1	0	0

*The Immigration Law and Regulatory Law subject areas were tested via performance test items only and therefore are not listed as subjects tested on the CBX exam.

Some commonalities were identified in the five race/ethnicity by subject matter analyses. For instance, the following subjects had at least one item flagged in the Non-Caucasian analysis and at

least one item was flagged in 50% of the other race/ethnicity by subject matter analyses: Contracts, Evidence, Professional Responsibilities, Real Property, Torts, and Criminal Law and Procedure.

Law School Type

A number of DIF analyses were conducted with respect to the law school type variable. In all cases, California ABA law schools were designated as the reference group and the focal groups for each analysis were California Accredited, California Registered, Out-of-state ABA, Attorney, and Foreign JD. The results of these analyses are presented in Table 12.

Table 12. Law School Type DIF Statistics

Focal Group	MH Statistic	LA-LOR Statistic					
		Favors Focal Group			Favors California ABA		
		Large	Moderate	Negligible	Negligible	Moderate	Large
California Accredited	Nonsig.	0	2	17	62	10	3
	Significant	1	0	0	8	17	32
California Registered	Nonsig.	0	0	6	26	20	2
	Significant	0	0	0	0	3	95
Out-of-state ABA	Nonsig.	0	0	9	47	0	0
	Significant	0	0	1	49	23	23
Attorney	Nonsig.	0	0	17	30	0	0
	Significant	6	0	2	26	33	44
Foreign JD	Nonsig.	0	0	6	21	9	1
	Significant	0	0	0	1	10	104

For each of the law school type DIF analyses presented in Table 12 more items were flagged for DIF in favor of the California ABA law schools. However, it is also important to note that the number of candidates from California ABA law schools tended to be much larger compared to the number of candidates from CA Accredited and California Registered law schools and the Attorney and Foreign JD categories (see Table 1). Large differences in sample size may have impacted the results, potentially resulting in flagging items for DIF at a higher or lower rate than if the groups had been of similar size.

A large proportion of items were flagged for DIF in the California Registered (98 or 64%), Attorney (83 or 55%), and Foreign JD (114 or 75%) analyses. Of the items flagged, only six favored the focal group after controlling for overall performance on the MBE.

A much smaller number of items were flagged for DIF in the California Accredited (50 or 33%) and Out-of-state ABA (46 or 30%) analyses. With the exception of one item, all flagged items favored California ABA candidates after controlling for overall performance on the MBE for the two aforementioned analyses.

The number of candidates from the largest groups, California ABA and Out-of-state ABA, tended to fluctuate during each administration window (more candidates in July compared to February). However, the number of candidates testing from the other categories tended to be both smaller and remain relatively stable during both administration windows (see Table 1). Therefore, an analysis of the number of items flagged during each administration window is included in Table 13.

Table 13. Law School and Administration Window DIF Statistics

		LA-LOR Statistic					
Admin Window	MH Statistic	Favors California Accredited			Favors California ABA		
		Large	Moderate	Negligible	Negligible	Moderate	Large
February	Nonsig.	0	2	14	34	10	3
	Significant	1	0	0	0	4	8
July	Nonsig.	0	0	3	28	0	0
	Significant	0	0	0	8	13	24

		LA-LOR Statistic					
Admin Window	MH Statistic	Favors California Registered			Favors California ABA		
		Large	Moderate	Negligible	Negligible	Moderate	Large
February	Nonsig.	0	0	6	20	15	2
	Significant	0	0	0	0	2	31
July	Nonsig.	0	0	0	6	5	0
	Significant	0	0	0	0	1	64

		LA-LOR Statistic					
Admin Window	MH Statistic	Favors Out-of-state ABA			Favors California ABA		
		Large	Moderate	Negligible	Negligible	Moderate	Large
February	Nonsig.	0	0	6	35	0	0
	Significant	0	0	0	4	13	18
July	Nonsig.	0	0	3	12	0	0
	Significant	0	0	1	45	10	5

		LA-LOR Statistic					
Admin Window	MH Statistic	Favors Attorney			Favors California ABA		
		Large	Moderate	Negligible	Negligible	Moderate	Large
February	Nonsig.	0	0	16	25	0	0
	Significant	6	0	2	14	8	11
July	Nonsig.	0	0	1	5	0	0
	Significant	0	0	0	12	25	33

		LA-LOR Statistic					
Admin Window	MH Statistic	Favors Foreign JD			Favors California ABA		
		Large	Moderate	Negligible	Negligible	Moderate	Large
February	Nonsig.	0	0	6	20	7	1
	Significant	0	0	0	0	4	38
July	Nonsig.	0	0	0	1	2	0
	Significant	0	0	0	1	6	66

The number of items flagged for DIF was higher in the July windows compared to the February windows for California Accredited (37 compared to 13, California Registered (65 compared to 33), Attorney (58 compared to 25, and Foreign JD (72 compared to 42 candidates). This may be a reflection of the large discrepancy between the number of candidates in each of these categories compared to the number of California ABA candidates in the July administration windows.

It is important to note that more items were flagged in the February administration window in the Out-of-state ABA (31 compared to 15) DIF analysis compared to the July window.

Gender, Race/Ethnicity, and Law School Type

Given that law school type DIF appeared to be flagged more often for items compared to gender or race/ethnicity DIF, additional analyses were conducted to investigate the relationship between all three demographic characteristics of CBX candidates.

It would be inappropriate to examine gender and race/ethnicity DIF within each law school type, due to limited sample sizes. For instance, for several law school types fewer than 50 candidates represented a race/ethnicity group over 20 administrations.

It is also impossible to disaggregate the gender and race/ethnicity DIF results from law school type, since candidates from all law school types were pooled in order to conduct each gender and race/ethnicity DIF analysis.

Therefore, the demographic characteristics of each law school type were examined to attempt to explain the relationship between gender and race/ethnicity with respect to law school type. Table 14 provides gender and race/ethnicity characteristics of each law school type. It is important to note that gender and race/ethnicity were not reported by all candidates for each law school type, therefore the sample sizes may not align with the values in Table 1.

Table 14. Gender, Race/Ethnicity, and Law School Type

Gender	Law School Type					
	CA ABA	CA Accredited	CA Registered	Out-of-state ABA	Attorney	Foreign JD
Male	15,921(50%)	976(52%)	546(59%)	5,945(55%)	2,929(53%)	1,074(43%)
Female	15,733(50%)	912(48%)	385(41%)	4,913(45%)	2,557(47%)	1,430(57%)
Race/ Ethnicity	CA ABA	CA Accredited	CA Registered	Out-of-state ABA	Attorney	Foreign JD
Asian	6,111 (20%)	226(12%)	114(12%)	2,230(22%)	1,126(21%)	1,340(54%)
Black/AA	941 (3%)	78 (4%)	48 (5%)	532 (5%)	260 (5%)	97 (4%)
Hispanic	3,260 (11%)	303(17%)	79 (9%)	872 (9%)	367 (7%)	164 (7%)
Caucasian	19,537(64%)	1,196(65%)	669(73%)	6,333(62%)	3,516(66%)	870(35%)
Other	488 (2%)	39 (2%)	12 (1%)	173 (2%)	65 (1%)	8(<1%)

From a gender perspective, the proportion of males and females was similar for CA ABA, CA Accredited, Out-of-state ABA, and Attorney law school types. The two least similar law school types with respect to gender were CA Registered, with a 59% to 41% male-to-female ratio, and Foreign JD, with a 43% to 57% male-to-female ratio.

From a race/ethnicity perspective, large proportional differences were observed for Asian candidates across law school types. Asian candidates represented 12% of all candidates for CA Accredited and CA Registered law school types and between 20% and 22% of all candidates for CA ABA, Out-of-state ABA, and Attorney law school types. However, Foreign JD law schools were represented by the largest proportion of Asian candidates at 54%.

Similarly, large proportional differences were also observed for Caucasian candidates for Foreign JD law schools compared to other law school types. While Caucasians represented between 62% and 73% of all candidates for all other law school types, Caucasians represented 35% of all candidates for Foreign JD law schools.

Relatively small proportional differences were observed for Hispanic candidates across law school types. Hispanic candidates represented between 7% and 11% of all candidates for all law school types, with the exception of CA Accredited law schools which had a slightly higher proportion of Hispanic candidates at 17%.

Only small proportional differences were observed for Black/African American candidates across all law school types. Black/African American candidates represented between 3% and 5% of all candidates for all law school types.

Similarly, small proportional differences were observed for candidates who identified as Other with respect to race/ethnicity across all law school types. Other candidates represented 2% or less of the proportion of all candidates for all law school types.

CONCLUSION

The purpose of this analysis was to examine DIF on the CBX essay and performance test items with respect to the following demographic variables: gender, race/ethnicity, and law school type. Trends related to item type, subject matter, and administration window were also of interest and are discussed in this section. In the final section, recommendations for the State Bar are presented.

Gender

The first demographic variable examined was gender. Female candidates performed better than male candidates on 99% of the CBX essay and performance test items after controlling for performance on the MBE. However, only 30 items, or 20%, were flagged for gender DIF based on the flagging criteria used in this study meaning gender DIF was considered negligible for 80% of items.

Since there appeared to be an overall bias favoring female candidates on the CBX, the rate at which items were flagged for gender DIF was investigated for each of the two item types. Both essay and performance test items had a similar proportion of items flagged for gender DIF. Therefore, it is possible that these results indicate a bias towards females for these two item types and the CBX overall. However, it is also possible that the results are confounded because of the choice to use the MBE, which consists of only multiple-choice items, as an indicator of overall performance in this study. Due to this limitation, conclusions should not be drawn about potential item type bias on the CBX based only on the results of this study. If multiple-choice data are available for the MBE, item type bias could be investigated in future studies.

Finally, it appears that some specific subjects were flagged for gender DIF at a higher rate compared to others. Specifically, Trusts (2 of 3, or 67%), Wills and Succession (4 of 7 items, or 57%), Regulatory Law (1 of 2, or 50%), and Community Property (5 of 11 items, or 45%) were flagged more often than other subject areas. On the other hand, no items from the following subjects were flagged for gender DIF: Constitutional Law, Criminal Law and Procedure, Immigration Law, Real Property, and Remedies. Although this finding may help to explain some of the gender DIF, the number of items administered for each subject varied greatly so these results should be interpreted with caution.

Race/Ethnicity

Four specific comparisons were of interest in the race/ethnicity DIF analysis. Caucasian candidates were designated as the reference variable and the focal groups for each analysis were Black/African American, Asian, Hispanic, or Other. All items flagged for race/ethnicity DIF favored Caucasians, although the proportion of items flagged were relatively small. Specifically, the largest proportion of items were flagged in the Black/African American DIF analysis (25 items or 16%), followed by the Asian (10 items or 7%), Hispanic (7 items or 5%), and Other (7 items or 5%) analyses.

It is important to note that large discrepancies between the number of candidates in the reference and focal groups were observed in each administration window, particularly in the Black/African American and Other analyses. It is difficult to draw conclusions given then large differences in sample sizes in each of the two aforementioned DIF analyses.

As a result, an additional variable (Non-Caucasian) was created by combining the Black/African American, Asian, Hispanic, and Other variables into a single variable. Fewer items were flagged for race/ethnicity DIF in the Non-Caucasian (8 items or 5%), analysis, although a similar trend was

observed. Generally, Caucasians performed better than Non-Caucasians after controlling for overall performance on the MBE, although observed DIF was considered negligible for 95% of items.

The Non-Caucasian variable was used as the primary variable for identifying DIF in follow-up studies, such as an analysis of DIF with respect to race/ethnicity and subject matter. In the analysis the following subjects had at least one item flagged in the Non-Caucasian analysis and were also flagged in a majority of the other race/ethnicity analyses: Contracts, Evidence, Professional Responsibilities, Real Property, Torts, and Criminal Law and Procedure. No other follow-up analyses yielded relationships worth noting.

Law School Type

The final DIF analysis centered around law school type. For this analysis California ABA law schools were designated as the reference group and the focal groups were California Accredited, California Registered, Out-of-state ABA, Attorney, and Foreign JD.

The law school type DIF analysis resulted in more items being flagged than in the gender and race/ethnicity analyses. A large proportion of items were flagged in the California Registered (98 or 64%), Attorney (83 or 55%), and Foreign JD (114 or 75%) analyses while smaller proportions of items were flagged in the California Accredited (50 or 33%) and Out-of-state ABA (46 or 30%) analyses. Generally, items flagged for DIF tended to favor California ABA candidates.

It is important to note that the number of candidates from the largest groups, California ABA and Out-of-state ABA, tended to fluctuate during each administration window, with more candidates testing in July compared to February. However, the number of candidates testing from the other categories tended to be both smaller and remain relatively stable during each annual administration window.

Upon further investigation it was discovered that the number of items flagged for DIF was higher in the July windows compared to the February windows for California Accredited, California Registered, Attorney, and Foreign JD candidates. On the other hand, more items were flagged in the February administration windows in the Out-of-state ABA DIF analysis.

There appears to be an administration window effect with respect to law school type DIF. The results may be attributed to large discrepancies between the number of reference and focal group candidates testing in a particular administration window, or it could be due to the time of year that particular students tend to graduate from each type of law school. Although it is difficult to assign a cause based only on the variables included in this study, it seems reasonable that candidates with different educational backgrounds would be expected to perform differently on items even after controlling for overall MBE score.

Gender, Race/Ethnicity, and Law School Type

Gender and race/ethnicity demographic variables were examined for each law school type to attempt to explain some of the DIF trends uncovered in this report.

From a gender perspective, the proportion of males and females was similar for CA ABA, CA Accredited, Out-of-state ABA, and Attorney law schools. The law school type with the largest proportion of males, at 59%, was CA Registered. The only law school type where females were the majority was the Foreign JD category, where 57% of all candidates were female.

From a race/ethnicity perspective, large proportional differences were observed for Asian and Caucasian candidates across law school types. The proportion of Asian candidates was between 12% and 22% for all law school types, with the exception of the Foreign JD category where 54% of all candidates were Asian. On the other hand, Caucasian candidates represented 35% of all candidates in the Foreign JD category, while Caucasian candidates represented between 62% and 73% of all candidates for all other law school types.

Relatively small proportional differences were observed for Hispanic candidates across law school types, with Hispanic candidates representing between 7% and 17% of all candidates across all law school types.

Only small proportional differences were observed for Black/African American candidates, who represented between 3% and 5% of all candidates across all law school types. Likewise, candidates who identified as Other with respect to race/ethnicity represented 2% or less of the proportion of all candidates across all law school types.

In summary, it appears that the demographic make-up of the Foreign JD law school type is unlike the other law school types included in this study with respect to gender and race/ethnicity. Foreign JD is the only law school type where females represent a majority with respect to gender and Asians represent a majority with respect to race/ethnicity. These demographic characteristics, coupled with a relatively small sample size of Foreign JD candidates, may help to explain why a large proportion of items were flagged for law school type DIF in the Foreign JD analysis.

All other law school types had somewhat comparable demographics with respect to gender and race/ethnicity, with only relatively small differences observed. It is important to note that these were the only two demographic variables included in this study, therefore other socio-economic variables may be better for explaining differences in the demographic characteristics of law schools.

Recommendations

In order to provide recommendations, it is important to first reflect on the overall performance of the CBX items for context.

The results of this DIF study indicate no major areas of concern for the CBX with respect to DIF. Only 5% of items were flagged in the race/ethnicity analysis based on the Caucasian/Non-Caucasian analysis, which is a small amount of DIF for an assessment. Although a large proportion of items were identified for law school type DIF, similar findings could be expected given that candidates have different educational backgrounds regarding legal topics.

A finding of potential concern was the moderate number of items flagged for DIF with respect to gender (20% of items flagged). It is difficult to explain why a general bias towards females was observed across almost all items, even when the level of DIF was negligible. The results may be influenced by using the MBE scaled score as an indicator of overall performance, which included only multiple-choice items, while conducting a DIF study on essay and performance test items. Although this is a limitation of this study, the MBE scaled score was chosen as an indicator of overall performance after reviewing existing DIF studies conducted on assessments with only a small number of polytomous items.

Considering the results, there are four ways the State Bar can use this study to improve the CBX moving forward.

First, it is recommended that the State Bar convene a panel of attorneys with varying demographic backgrounds to conduct a bias and sensitivity review of the flagged items. Reviewers should look for particular words and phrases that may be biased towards a particular gender or race/ethnicity or misinterpreted depending on a candidate's background. Since all items included in this study have been administered, and thus will not appear on future forms, any comments from the reviewers should be used to inform future item development.

In addition to the initial review, bias and sensitivity reviews should take place for items appearing on future forms. These reviews should be incorporated into the examination review process prior to the administration of new forms.

Third, it is recommended that the State Bar use these results to inform the future design of the bar examination. When considering which item types are best for assessing legal knowledge it is also important to consider potential item type bias.

The final recommendation for the State Bar is to review the CBX scoring process. Although scoring was not considered in this study, there is always potential for bias due to the nature in which essays and performance tests are scored. Any rubrics or methods for choosing, training, or assigning raters should be reviewed.

REFERENCES

- Penfield, Randall (2013). DIFAS 5.0 Differential item functioning analysis system user's manual.
- Penfield, R. D., & Algina, J. (2006). A generalized DIF effect variance estimator for measuring unsigned differential test functioning in mixed format tests. *Journal of Educational Measurement*, 43(4), 295–312. <https://doi.org/10.1111/j.1745-3984.2006.00018.x>
- SAS Institute (2014). SAS version 9.4. Cary, NC.
- Zieky, M. (1993). Practical questions in the use of DIF statistics in item development. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337–347). Hillsdale, NJ: Lawrence Erlbaum.

APPENDIX A: GENDER ITEM-LEVEL DIF RESULTS

Note: Bolded cells correspond to statistically significant MH statistics and LA-LOR statistics with an absolute value greater than 0.43.

Reference		Male	
Focal		Female	
Year	Item	MH Statistic	LA-LOR
G0907	Q1	16.81	-0.22
G0907	Q2	1.46	-0.07
G0907	Q3	18.43	-0.24
G0907	Q4	1.80	0.07
G0907	Q5	41.20	-0.34
G0907	Q6	26.26	-0.28
G0907	Q7	38.35	-0.34
G0907	Q8	34.73	-0.32
G1002	Q1	12.84	-0.46
G1002	Q2	1.07	-0.14
G1002	Q3	15.50	-0.50
G1002	Q4	3.95	-0.25
G1002	Q5	0.36	-0.08
G1002	Q6	2.36	-0.19
G1002	Q7	8.31	-0.36
G1002	Q8	7.13	-0.33
G1007	Q1	25.30	-0.24
G1007	Q2	147.10	-0.60
G1007	Q3	79.02	-0.43
G1007	Q4	26.08	-0.25
G1007	Q5	0.76	-0.04
G1007	Q6	59.93	-0.37
G1007	Q7	83.40	-0.43
G1007	Q8	89.33	-0.45
G1102	Q1	17.05	-0.52
G1102	Q2	0.30	-0.07
G1102	Q3	0.14	-0.05
G1102	Q4	0.20	-0.06
G1102	Q5	13.65	-0.47
G1102	Q6	5.11	-0.29
G1102	Q7	14.36	-0.47
G1102	Q8	6.68	-0.32
G1107	Q1	9.41	-0.17
G1107	Q2	30.19	-0.31
G1107	Q3	5.10	-0.13

Reference		Male	
Focal		Female	
Year	Item	MH Statistic	LA-LOR
G1402	Q5	8.93	-0.35
G1402	Q6	1.33	-0.14
G1402	Q7	9.47	-0.36
G1402	Q8	5.68	-0.28
G1407	Q1	18.77	-0.25
G1407	Q2	15.76	-0.22
G1407	Q3	10.14	-0.19
G1407	Q4	37.29	-0.35
G1407	Q5	38.19	-0.36
G1407	Q6	20.77	-0.26
G1407	Q7	18.32	-0.25
G1407	Q8	42.24	-0.37
G1502	Q1	4.83	-0.27
G1502	Q2	4.61	-0.26
G1502	Q3	20.30	-0.57
G1502	Q4	2.22	-0.19
G1502	Q5	9.18	-0.38
G1502	Q6	32.78	-0.73
G1502	Q7	6.10	-0.30
G1502	Q8	4.31	-0.25
G1507	Q1	62.99	-0.48
G1507	Q2	27.70	-0.31
G1507	Q3	0.40	-0.04
G1507	Q4	77.84	-0.53
G1507	Q5	29.49	-0.33
G1507	Q6	0.09	-0.02
G1507	Q7	33.27	-0.34
G1507	Q8	62.35	-0.47
G1602	Q1	13.27	-0.49
G1602	Q2	0.59	-0.10
G1602	Q3	9.14	-0.41
G1602	Q4	5.94	-0.33
G1602	Q5	3.18	-0.24
G1602	Q6	2.06	-0.19
G1602	Q7	11.74	-0.48

Reference		Male	
Focal		Female	
Year	Item	MH Statistic	LA- LOR
G1107	Q4	1.94	-0.08
G1107	Q5	4.47	-0.12
G1107	Q6	65.93	-0.45
G1107	Q7	24.15	-0.27
G1107	Q8	48.35	-0.38
G1202	Q1	18.65	-0.53
G1202	Q2	2.93	-0.22
G1202	Q3	8.16	-0.35
G1202	Q4	4.28	-0.25
G1202	Q5	6.45	-0.32
G1202	Q6	6.75	-0.32
G1202	Q7	17.41	-0.52
G1202	Q8	5.51	-0.28
G1207	Q1	34.06	-0.31
G1207	Q2	44.14	-0.37
G1207	Q3	89.73	-0.51
G1207	Q4	12.87	-0.20
G1207	Q5	42.25	-0.35
G1207	Q6	26.04	-0.28
G1207	Q7	31.46	-0.30
G1207	Q8	110.77	-0.56
G1302	Q1	1.11	-0.13
G1302	Q2	4.23	-0.26
G1302	Q3	0.79	-0.11
G1302	Q4	6.35	-0.32
G1302	Q5	0.00	0.00
G1302	Q6	0.79	-0.11
G1302	Q7	10.33	-0.41
G1302	Q8	4.73	-0.27
G1307	Q1	32.86	-0.31
G1307	Q2	53.50	-0.39
G1307	Q3	28.83	-0.29
G1307	Q4	14.18	-0.20
G1307	Q5	34.60	-0.32
G1307	Q6	0.00	0.00
G1307	Q7	6.34	-0.14
G1307	Q8	66.84	-0.44
G1402	Q1	0.95	-0.12
G1402	Q2	14.77	-0.46
G1402	Q3	0.11	-0.04

Reference		Male	
Focal		Female	
Year	Item	MH Statistic	LA- LOR
G1602	Q8	14.13	-0.51
G1607	Q1	12.58	-0.24
G1607	Q2	10.39	-0.22
G1607	Q3	3.81	-0.13
G1607	Q4	11.65	-0.23
G1607	Q5	50.93	-0.47
G1607	Q6	45.66	-0.44
G1607	Q7	9.40	-0.20
G1607	Q8	31.20	-0.37
G1702	Q1	2.07	-0.22
G1702	Q2	0.27	-0.08
G1702	Q3	8.93	-0.47
G1702	Q4	2.24	-0.23
G1702	Q5	4.27	-0.32
G1702	Q6	3.66	-0.29
G1702	Q7	0.66	-0.12
G1702	Q8	1.43	-0.18
G1707	Q1	62.78	-0.48
G1707	Q2	44.97	-0.40
G1707	Q3	22.75	-0.29
G1707	Q4	18.70	-0.26
G1707	Q5	8.31	-0.17
G1707	Q7	32.16	-0.33
G1802	Q1	1.96	-0.19
G1802	Q2	0.36	-0.08
G1802	Q3	0.35	-0.08
G1802	Q4	2.75	-0.22
G1802	Q5	21.83	-0.63
G1802	Q7	10.38	-0.43
G1807	Q1	19.97	-0.24
G1807	Q2	70.81	-0.45
G1807	Q3	116.47	-0.58
G1807	Q4	47.34	-0.36
G1807	Q5	27.54	-0.29
G1807	Q7	54.52	-0.39
G1902	Q1	11.64	-0.42
G1902	Q2	8.07	-0.36
G1902	Q3	6.16	-0.31
G1902	Q4	10.00	-0.41
G1902	Q5	0.31	-0.07

Reference		Male	
Focal		Female	
Year	Item	MH Statistic	LA- LOR
G1402	Q4	0.09	-0.04

Reference		Male	
Focal		Female	
Year	Item	MH Statistic	LA- LOR
G1902	Q7	0.12	-0.04

APPENDIX B: RACE/ETHNICITY ITEM-LEVEL DIF RESULTS

Note: Bolded cells correspond to statistically significant MH statistics and LA-LOR statistics with an absolute value greater than 0.43.

Reference		Caucasian									
Focal		Non-Caucasian		Black/AA		Asian		Hispanic		Other	
Year	Item	MH Statistic	LA-LOR	MH Statistic	LA-LOR	MH Statistic	LA-LOR	MH Statistic	LA-LOR	MH Statistic	LA-LOR
G0907	Q1	0.37	-0.04	0.86	-0.15	0.26	-0.04	0.40	-0.07	2.03	0.38
G0907	Q2	0.26	0.03	0.36	0.10	0.19	0.03	0.61	-0.08	2.22	0.39
G0907	Q3	0.48	0.04	11.93	0.63	2.48	-0.11	2.82	0.19	2.77	0.41
G0907	Q4	3.74	0.12	0.42	0.11	4.58	0.15	0.02	-0.02	0.79	0.22
G0907	Q5	1.96	-0.08	2.85	0.29	12.28	-0.24	1.34	0.12	2.43	0.37
G0907	Q6	0.35	-0.04	3.00	0.30	0.33	-0.04	1.61	-0.14	0.02	-0.04
G0907	Q7	3.98	-0.12	6.04	0.42	13.16	-0.25	0.02	0.02	1.18	0.24
G0907	Q8	0.13	-0.02	5.56	0.42	1.56	-0.09	0.75	-0.09	4.58	0.48
G1002	Q1	0.55	0.11	0.06	0.11	1.34	0.19	0.02	-0.04	0.87	-0.53
G1002	Q2	0.37	0.09	0.04	-0.07	1.01	0.18	0.11	-0.08	0.13	0.22
G1002	Q3	0.92	0.14	0.09	0.11	0.44	0.11	1.54	0.33	0.84	-0.63
G1002	Q4	1.15	0.15	2.39	0.65	0.08	0.04	1.91	0.36	0.01	-0.04
G1002	Q5	0.07	-0.04	0.61	0.33	0.03	0.03	0.68	-0.22	1.20	-0.57
G1002	Q6	0.00	0.00	0.01	0.05	0.59	-0.13	1.31	0.31	0.59	0.49
G1002	Q7	4.14	0.28	1.33	0.49	2.21	0.24	1.97	0.35	0.07	-0.13
G1002	Q8	1.04	-0.14	1.35	-0.49	0.76	-0.14	0.11	0.08	1.53	-0.63
G1007	Q1	1.04	0.05	7.06	0.37	0.01	0.01	0.19	0.04	0.18	0.08
G1007	Q2	0.37	-0.03	3.00	0.25	0.60	-0.05	1.69	-0.12	0.14	0.08
G1007	Q3	0.21	0.02	12.74	0.52	2.04	-0.09	1.58	0.11	1.53	0.25
G1007	Q4	5.95	0.13	4.79	0.30	2.17	0.09	2.19	0.13	0.81	0.18
G1007	Q5	15.55	0.21	5.36	0.32	20.96	0.28	0.00	0.00	0.18	0.08
G1007	Q6	0.47	0.04	7.69	0.40	0.00	0.00	0.10	0.03	0.27	-0.11
G1007	Q7	0.37	0.03	27.51	0.73	4.65	-0.13	5.34	0.20	0.09	-0.06
G1007	Q8	0.01	-0.01	16.19	0.55	4.93	-0.13	2.64	0.14	0.51	-0.14
G1102	Q1	4.68	0.30	8.62	1.42	1.37	0.19	0.71	0.21	0.57	0.34
G1102	Q2	0.35	0.08	0.01	0.03	1.26	0.18	1.47	-0.30	0.77	0.43
G1102	Q3	0.24	0.07	0.12	0.13	0.30	-0.09	5.18	0.59	0.09	-0.16
G1102	Q4	0.53	0.10	0.02	0.05	0.14	0.06	0.85	0.24	0.02	-0.06
G1102	Q5	0.02	-0.02	0.60	0.30	0.12	-0.06	0.27	-0.14	0.19	0.19
G1102	Q6	0.91	0.13	1.06	0.38	0.93	0.16	0.12	-0.09	0.14	0.17
G1102	Q7	0.01	0.01	0.38	0.24	0.47	-0.11	1.25	0.28	0.11	0.16
G1102	Q8	0.40	0.09	2.45	0.60	0.03	0.03	0.03	0.05	0.30	0.29
G1107	Q1	8.04	0.17	0.07	0.05	5.87	0.18	3.37	0.20	2.48	0.36
G1107	Q2	0.24	-0.03	1.84	0.24	2.09	-0.11	0.26	0.05	0.08	-0.06
G1107	Q3	4.53	0.13	16.16	0.67	1.40	0.09	1.48	0.13	2.05	-0.33

Reference		Caucasian									
Focal		Non-Caucasian		Black/AA		Asian		Hispanic		Other	
Year	Item	MH Statistic	LA-LOR	MH Statistic	LA-LOR	MH Statistic	LA-LOR	MH Statistic	LA-LOR	MH Statistic	LA-LOR
G1107	Q4	11.86	0.21	1.16	0.19	8.76	0.22	4.10	0.22	1.52	0.27
G1107	Q5	1.85	0.08	0.85	-0.15	0.41	0.05	8.23	0.30	0.15	-0.09
G1107	Q6	0.18	0.03	3.37	0.32	0.10	0.02	2.43	-0.16	3.23	0.42
G1107	Q7	0.95	-0.06	1.76	0.21	5.45	-0.17	0.16	0.04	0.79	0.19
G1107	Q8	0.02	0.01	2.03	0.24	0.47	-0.05	0.38	0.06	0.00	-0.01
G1202	Q1	0.47	-0.09	0.17	0.14	0.14	-0.06	3.25	-0.42	1.44	0.66
G1202	Q2	1.97	-0.19	1.44	0.37	2.31	-0.25	3.00	-0.39	0.09	-0.15
G1202	Q3	0.19	-0.06	0.24	0.14	0.91	-0.15	0.24	-0.12	2.10	0.75
G1202	Q4	0.16	-0.05	2.02	0.44	1.52	-0.20	0.30	0.13	0.56	-0.43
G1202	Q5	3.23	0.24	0.08	-0.09	5.03	0.37	0.03	-0.04	3.49	1.10
G1202	Q6	0.64	-0.11	2.51	0.58	3.11	-0.30	0.20	-0.12	0.71	0.46
G1202	Q7	0.06	0.03	0.02	0.04	0.31	0.09	0.00	-0.01	0.61	-0.47
G1202	Q8	1.22	0.15	0.66	0.26	0.93	0.16	0.00	0.00	1.98	0.77
G1207	Q1	1.67	0.08	7.02	0.43	0.00	0.00	1.73	0.13	0.14	-0.11
G1207	Q2	0.01	-0.01	0.26	0.08	0.75	0.06	3.46	-0.19	0.09	-0.09
G1207	Q3	1.15	0.06	6.33	0.42	0.27	-0.04	2.77	0.17	0.09	0.09
G1207	Q4	8.29	0.17	6.71	0.44	2.08	0.10	5.80	0.25	0.11	0.10
G1207	Q5	3.42	-0.11	0.67	0.14	9.63	-0.21	0.41	0.07	0.15	-0.11
G1207	Q6	0.07	-0.02	7.07	0.44	0.00	0.00	4.82	-0.22	0.08	0.08
G1207	Q7	1.06	-0.06	4.56	0.34	6.26	-0.17	0.90	0.10	0.09	-0.08
G1207	Q8	5.17	-0.13	1.81	0.22	18.96	-0.31	0.77	0.09	0.49	0.19
G1302	Q1	0.50	0.09	0.01	-0.03	0.50	0.11	0.61	0.20	0.01	0.07
G1302	Q2	0.08	0.04	2.59	0.56	0.02	-0.03	0.01	0.03	0.13	-0.18
G1302	Q3	0.64	0.11	4.16	0.72	0.01	-0.02	0.02	0.04	0.76	0.58
G1302	Q4	0.67	-0.11	0.18	0.13	2.17	-0.24	0.08	-0.07	2.71	0.96
G1302	Q5	3.77	0.27	0.65	0.23	0.24	0.08	7.10	0.73	1.80	0.90
G1302	Q6	2.23	0.21	1.68	0.44	0.99	0.16	0.77	0.23	0.04	-0.10
G1302	Q7	3.61	0.26	2.06	0.38	0.71	0.14	4.54	0.58	0.06	0.10
G1302	Q8	0.00	0.00	0.79	0.26	0.02	0.02	0.25	-0.13	1.09	-0.46
G1307	Q1	6.74	0.15	5.79	0.44	9.54	0.22	0.07	-0.03	0.00	0.01
G1307	Q2	5.09	0.13	7.07	0.45	0.21	0.03	5.44	0.23	3.24	0.40
G1307	Q3	0.36	0.04	1.24	0.20	0.00	0.00	0.00	0.00	5.40	0.52
G1307	Q4	2.58	0.09	3.99	0.36	0.08	-0.02	5.82	0.24	4.82	0.46
G1307	Q5	1.62	0.08	16.45	0.76	0.23	0.03	0.02	0.02	0.04	0.04
G1307	Q6	9.46	0.18	4.91	0.39	4.34	0.15	4.43	0.21	1.00	0.22
G1307	Q7	0.77	0.05	2.65	0.29	0.07	-0.02	2.26	0.15	0.27	0.11
G1307	Q8	0.14	0.02	6.99	0.45	0.52	-0.05	0.59	0.08	0.00	-0.01
G1402	Q1	6.87	0.37	0.70	0.30	7.97	0.47	0.00	-0.01	5.49	1.91
G1402	Q2	2.70	0.23	0.41	0.23	3.75	0.32	0.18	-0.10	0.75	0.53
G1402	Q3	4.51	0.29	5.62	0.79	2.54	0.26	0.00	0.01	0.16	0.23

Reference		Caucasian									
Focal		Non-Caucasian		Black/AA		Asian		Hispanic		Other	
Year	Item	MH Statistic	LA- LOR								
G1402	Q4	16.11	0.58	0.11	0.11	12.24	0.61	5.93	0.66	1.82	0.95
G1402	Q5	0.34	0.08	1.40	0.40	0.40	0.10	0.19	-0.10	0.12	0.21
G1402	Q6	1.43	0.16	1.77	0.51	2.42	0.25	0.37	-0.14	0.00	0.02
G1402	Q7	3.23	0.24	0.05	0.07	2.24	0.24	2.70	0.37	0.13	-0.21
G1402	Q8	5.12	0.30	0.05	0.07	1.47	0.19	4.98	0.56	1.55	0.83
G1407	Q1	8.79	0.18	2.03	0.22	4.19	0.15	5.53	0.25	0.82	0.22
G1407	Q2	1.48	0.07	4.75	0.34	0.20	-0.03	3.02	0.18	0.50	0.17
G1407	Q3	0.89	0.06	2.03	0.24	0.04	-0.02	1.50	0.13	2.40	0.39
G1407	Q4	7.07	0.16	5.85	0.38	4.73	0.16	0.66	0.08	0.03	0.04
G1407	Q5	2.13	0.09	4.24	0.33	0.49	-0.05	7.96	0.30	1.62	0.30
G1407	Q6	8.14	0.17	12.61	0.55	1.31	0.08	3.10	0.18	2.03	0.34
G1407	Q7	5.88	0.15	16.23	0.68	0.00	0.00	11.05	0.35	0.19	-0.10
G1407	Q8	1.86	0.08	5.12	0.34	0.08	-0.02	4.33	0.21	0.68	0.18
G1502	Q1	0.90	-0.13	0.38	-0.21	0.00	-0.01	3.81	-0.42	0.14	-0.15
G1502	Q2	3.72	-0.26	1.34	-0.34	3.90	-0.33	0.52	-0.16	0.33	-0.23
G1502	Q3	0.06	-0.03	0.06	-0.07	0.13	-0.06	0.52	0.16	1.77	-0.55
G1502	Q4	0.20	0.06	2.82	0.56	0.40	-0.11	1.39	0.27	0.01	-0.04
G1502	Q5	0.11	-0.05	0.00	0.00	1.54	-0.20	2.65	0.38	0.99	-0.38
G1502	Q6	0.07	0.04	1.41	0.40	0.08	0.05	0.02	0.03	3.33	-0.87
G1502	Q7	0.16	-0.06	0.01	0.03	0.50	-0.12	0.00	0.00	0.01	-0.04
G1502	Q8	0.09	0.04	0.42	0.18	0.02	0.02	0.02	-0.03	0.02	0.06
G1507	Q1	0.41	-0.04	0.32	0.09	1.86	-0.10	0.08	0.03	0.03	-0.04
G1507	Q2	1.66	0.08	3.02	0.28	0.58	0.06	0.28	0.05	0.10	0.08
G1507	Q3	5.70	0.15	1.51	0.20	11.80	0.26	1.74	-0.13	0.58	0.21
G1507	Q4	0.00	0.00	9.03	0.50	2.48	-0.12	0.00	0.00	2.75	0.43
G1507	Q5	0.00	0.00	4.85	0.36	2.24	-0.11	0.25	0.05	1.86	0.38
G1507	Q6	0.58	0.05	3.55	0.32	1.00	0.08	1.67	-0.13	1.22	0.28
G1507	Q7	0.13	-0.02	11.65	0.57	5.93	-0.18	0.72	0.08	0.91	0.24
G1507	Q8	2.96	0.11	6.38	0.42	0.56	0.06	0.63	0.08	0.97	0.25
G1602	Q1	0.03	0.02	1.62	0.45	0.00	0.00	0.03	0.04	0.19	-0.20
G1602	Q2	9.54	0.44	2.76	0.59	9.73	0.54	2.00	0.35	1.52	-0.60
G1602	Q3	1.04	0.15	0.62	0.28	2.04	0.24	0.11	-0.08	0.19	0.19
G1602	Q4	1.06	0.15	0.30	-0.18	2.10	0.25	0.00	0.02	1.79	0.76
G1602	Q5	0.21	0.07	0.14	-0.13	2.37	0.26	0.86	-0.24	0.67	-0.40
G1602	Q6	1.54	0.18	1.23	0.36	2.99	0.29	0.23	-0.12	0.05	0.11
G1602	Q7	0.94	0.14	2.16	0.51	0.34	0.10	0.86	0.24	0.21	-0.20
G1602	Q8	1.46	0.17	2.73	0.59	1.10	0.17	0.13	0.09	0.51	0.38
G1607	Q1	1.71	-0.09	0.85	0.17	1.65	-0.11	1.40	-0.13	1.04	-0.25
G1607	Q2	0.15	0.03	7.74	0.53	0.13	0.03	0.43	-0.07	0.59	-0.19
G1607	Q3	3.92	0.14	8.23	0.56	0.00	0.00	5.75	0.27	1.22	0.31

Reference		Caucasian									
Focal		Non-Caucasian		Black/AA		Asian		Hispanic		Other	
Year	Item	MH Statistic	LA-LOR	MH Statistic	LA-LOR	MH Statistic	LA-LOR	MH Statistic	LA-LOR	MH Statistic	LA-LOR
G1607	Q4	0.07	0.02	6.91	0.50	0.00	0.00	1.37	-0.13	2.13	0.39
G1607	Q5	0.31	0.04	0.68	0.16	0.37	0.05	0.12	-0.04	0.55	0.18
G1607	Q6	2.98	-0.12	0.03	0.03	2.43	-0.13	2.89	-0.18	0.20	0.12
G1607	Q7	1.62	0.09	11.18	0.67	0.50	-0.06	3.30	0.20	0.02	0.04
G1607	Q8	0.88	-0.07	11.97	0.66	10.01	-0.27	0.59	0.09	0.20	0.11
G1702	Q1	1.27	0.17	0.46	0.24	3.27	0.32	1.25	-0.26	0.86	0.45
G1702	Q2	3.15	0.28	0.04	0.06	10.90	0.69	1.76	-0.30	0.06	0.16
G1702	Q3	0.17	0.07	0.89	0.33	0.54	0.14	1.73	-0.32	1.27	0.76
G1702	Q4	0.79	-0.14	0.11	0.11	0.70	-0.17	0.96	-0.25	0.02	-0.07
G1702	Q5	0.55	0.12	0.20	0.15	1.75	0.26	0.21	-0.11	0.15	-0.26
G1702	Q6	12.67	0.56	1.74	0.47	16.15	0.78	1.19	0.29	0.15	0.21
G1702	Q7	0.69	-0.13	0.10	0.11	1.15	-0.20	0.00	0.01	1.74	-0.74
G1702	Q8	0.07	0.04	1.22	0.38	0.01	-0.02	0.11	-0.08	1.77	0.77
G1707	Q1	2.96	0.11	4.12	0.32	7.08	0.20	2.47	-0.15	0.12	0.09
G1707	Q2	1.56	0.08	2.04	0.21	1.38	0.09	0.01	0.01	0.38	-0.14
G1707	Q3	6.31	0.16	3.51	0.29	5.44	0.18	0.11	0.03	1.80	0.31
G1707	Q4	4.35	0.13	0.45	0.10	7.24	0.20	0.06	0.02	0.13	-0.09
G1707	Q5	3.00	0.11	4.45	0.32	3.61	0.14	0.01	-0.01	0.01	0.02
G1707	Q7	20.59	0.27	1.43	0.17	18.13	0.31	7.18	0.26	0.20	0.10
G1802	Q1	14.83	0.55	2.24	0.49	6.87	0.47	11.03	0.78	0.86	0.44
G1802	Q2	0.63	0.11	0.47	0.22	2.99	0.31	0.61	-0.18	0.34	-0.24
G1802	Q3	1.92	0.19	0.39	0.17	2.72	0.27	0.03	0.04	0.00	0.01
G1802	Q4	5.41	0.33	2.81	0.51	3.77	0.34	1.98	0.33	0.56	-0.33
G1802	Q5	0.05	0.03	3.37	0.57	0.04	-0.04	0.00	0.01	0.14	-0.18
G1802	Q7	0.26	-0.07	5.77	0.78	3.75	-0.32	0.00	0.01	2.18	0.67
G1807	Q1	0.78	0.05	3.94	0.25	0.50	0.05	0.20	-0.04	0.78	0.18
G1807	Q2	21.17	0.25	3.51	0.23	24.51	0.33	1.85	0.11	4.28	0.44
G1807	Q3	17.30	0.23	6.02	0.30	18.08	0.28	0.78	0.07	9.52	0.65
G1807	Q4	2.86	0.09	4.01	0.24	7.61	0.18	1.29	-0.09	0.09	0.06
G1807	Q5	1.13	0.06	3.45	0.24	0.59	0.05	0.10	-0.03	4.80	0.47
G1807	Q7	5.03	0.12	1.46	0.14	5.68	0.15	0.04	0.02	2.38	0.30
G1902	Q1	2.59	0.20	3.12	0.53	2.11	0.21	0.10	0.06	0.04	0.15
G1902	Q2	18.89	0.56	2.33	0.45	27.77	0.79	1.08	0.20	0.01	-0.09
G1902	Q3	20.88	0.58	3.79	0.60	21.36	0.69	5.07	0.43	1.72	0.98
G1902	Q4	20.88	0.61	2.32	0.46	19.63	0.69	4.12	0.41	0.99	0.86
G1902	Q5	15.50	0.52	3.52	0.58	24.86	0.78	0.00	0.00	0.70	0.64
G1902	Q7	3.58	0.24	2.65	0.48	3.01	0.26	0.92	0.20	0.07	-0.17

APPENDIX C: LAW SCHOOL TYPE ITEM-LEVEL DIF RESULTS

Note: Bolded cells correspond to statistically significant MH statistics and LA-LOR statistics with an absolute value greater than 0.43.

Reference		CA ABA									
Focal		CA Accredited		CA Registered		Out-of-state ABA		Attorney		Foreign JD	
Year	Item	MH Statistic	LA-LOR	MH Statistic	LA-LOR	MH Statistic	LA-LOR	MH Statistic	LA-LOR	MH Statistic	LA-LOR
G0907	Q1	3.65	0.32	9.02	0.78	23.30	0.32	27.45	0.68	2.84	0.44
G0907	Q2	17.34	0.74	25.59	1.41	4.60	0.14	25.38	0.66	22.92	1.34
G0907	Q3	1.86	0.22	9.57	0.87	8.99	0.20	11.73	0.44	13.77	0.98
G0907	Q4	14.77	0.63	14.36	1.08	0.22	0.03	20.36	0.59	21.14	1.27
G0907	Q5	34.53	0.94	71.56	2.16	26.63	0.32	84.63	1.09	8.89	0.73
G0907	Q6	2.54	0.26	14.08	1.07	27.22	0.34	44.91	0.86	3.14	0.45
G0907	Q7	16.29	0.65	23.65	1.27	5.12	0.15	31.11	0.70	14.00	0.92
G0907	Q8	19.80	0.73	24.78	1.27	1.79	0.09	10.25	0.38	24.92	1.35
G1002	Q1	2.19	0.48	1.46	0.38	0.98	0.19	1.86	0.21	0.40	0.24
G1002	Q2	5.21	0.80	2.06	0.43	0.79	0.17	4.69	0.34	0.01	0.04
G1002	Q3	4.59	0.65	0.26	0.14	2.18	0.28	0.82	0.14	0.02	-0.04
G1002	Q4	0.21	0.13	1.77	0.39	9.44	0.57	22.30	0.73	4.94	0.77
G1002	Q5	1.00	-0.31	5.74	0.77	1.21	0.21	14.91	0.62	3.66	0.63
G1002	Q6	1.82	0.41	3.45	0.57	8.52	0.55	24.59	0.75	3.37	0.65
G1002	Q7	1.25	0.32	1.91	0.38	3.84	0.37	5.27	0.35	9.40	1.04
G1002	Q8	6.51	0.84	8.91	0.88	6.46	0.49	4.39	0.31	8.73	0.95
G1007	Q1	1.84	0.16	19.80	0.85	49.01	0.40	65.10	0.82	62.57	1.30
G1007	Q2	6.54	0.32	14.11	0.77	45.12	0.40	36.29	0.67	38.63	1.16
G1007	Q3	10.36	0.38	43.87	1.27	46.30	0.40	24.85	0.53	50.16	1.14
G1007	Q4	10.29	0.39	30.42	1.10	6.44	0.15	9.04	0.31	32.49	1.00
G1007	Q5	6.90	0.32	22.05	0.96	46.38	0.40	38.69	0.65	44.99	1.12
G1007	Q6	0.05	0.03	19.71	0.94	48.98	0.42	27.70	0.54	7.02	0.44
G1007	Q7	26.07	0.61	67.15	1.61	7.38	0.16	28.41	0.53	54.36	1.16
G1007	Q8	46.18	0.82	51.55	1.32	0.30	-0.03	1.58	0.12	13.76	0.58
G1102	Q1	0.28	0.19	10.82	1.08	1.25	0.21	0.12	0.05	3.26	0.54
G1102	Q2	0.04	0.07	19.26	1.34	4.94	0.40	29.49	0.81	20.95	1.43
G1102	Q3	0.32	0.19	5.00	0.67	11.80	0.65	14.99	0.58	9.18	0.93
G1102	Q4	0.11	-0.11	11.52	1.02	0.11	0.06	1.07	0.16	14.16	1.13
G1102	Q5	2.98	0.65	13.05	1.17	0.07	0.05	1.51	0.19	0.83	0.27
G1102	Q6	1.16	-0.35	6.29	0.82	0.33	0.11	1.24	0.17	4.45	0.67
G1102	Q7	4.12	0.60	40.07	2.15	1.54	-0.22	2.28	-0.22	8.33	0.88
G1102	Q8	1.96	0.46	15.78	1.26	2.64	0.30	4.67	0.32	9.90	0.91
G1107	Q1	12.03	0.63	5.56	0.66	36.53	0.41	33.51	0.71	35.26	1.48
G1107	Q2	10.51	0.64	37.12	2.34	62.17	0.55	24.28	0.60	31.29	1.57
G1107	Q3	5.18	0.46	3.71	0.57	46.20	0.47	40.06	0.78	27.68	1.37

Reference		CA ABA									
Focal		CA Accredited		CA Registered		Out-of-state ABA		Attorney		Foreign JD	
Year	Item	MH Statistic	LA- LOR	MH Statistic	LA- LOR	MH Statistic	LA- LOR	MH Statistic	LA- LOR	MH Statistic	LA- LOR
G1107	Q4	10.10	0.65	1.57	0.39	5.42	0.16	16.86	0.51	4.42	0.52
G1107	Q5	2.82	0.30	16.36	1.29	54.24	0.49	12.80	0.45	5.16	0.54
G1107	Q6	0.09	0.05	2.12	0.44	40.93	0.44	48.20	0.92	9.24	0.79
G1107	Q7	28.33	1.02	31.92	1.95	0.70	0.06	5.38	0.27	24.87	1.26
G1107	Q8	30.25	1.12	21.75	1.57	0.32	0.04	18.80	0.53	26.91	1.36
G1202	Q1	0.44	-0.20	8.57	0.89	8.44	0.52	11.85	0.51	4.03	0.62
G1202	Q2	0.03	0.06	9.63	0.95	10.47	0.58	6.96	0.41	0.06	0.08
G1202	Q3	1.20	-0.32	3.02	0.52	14.48	0.66	7.89	0.42	1.28	0.30
G1202	Q4	1.73	0.43	1.94	0.41	3.91	0.36	6.13	0.38	0.94	0.28
G1202	Q5	1.43	0.36	1.06	0.31	0.28	0.09	0.54	0.11	0.72	0.27
G1202	Q6	0.27	0.17	4.12	0.62	11.92	0.65	3.03	0.27	1.05	0.34
G1202	Q7	1.73	0.43	4.23	0.64	4.42	0.38	0.02	-0.02	8.42	0.87
G1202	Q8	0.00	0.02	13.07	1.08	3.48	0.33	4.41	0.31	7.25	0.79
G1207	Q1	12.98	0.63	17.82	1.20	11.59	0.22	26.92	0.57	37.57	1.43
G1207	Q2	1.01	0.18	6.41	0.81	21.85	0.32	18.89	0.51	10.31	0.74
G1207	Q3	5.60	0.45	15.59	1.29	38.02	0.41	38.79	0.70	42.35	1.68
G1207	Q4	1.12	0.19	14.20	1.32	2.05	0.10	5.98	0.29	16.45	1.01
G1207	Q5	9.03	0.56	7.27	0.87	30.47	0.36	51.50	0.83	15.66	0.93
G1207	Q6	3.93	0.34	11.35	1.14	35.21	0.40	64.77	0.94	65.43	2.24
G1207	Q7	14.57	0.72	5.85	0.74	1.83	0.09	0.72	0.10	18.31	1.01
G1207	Q8	35.14	1.15	32.50	1.73	23.31	0.31	39.63	0.69	41.57	1.43
G1302	Q1	0.24	0.18	0.14	-0.12	3.55	-0.34	7.31	-0.41	2.17	0.40
G1302	Q2	5.61	0.92	12.57	1.28	0.90	0.18	0.03	0.03	0.00	0.02
G1302	Q3	0.01	0.04	9.88	1.08	29.48	1.08	24.82	0.78	2.55	0.43
G1302	Q4	1.84	0.43	8.88	0.87	19.11	0.80	28.19	0.81	8.08	0.79
G1302	Q5	1.64	0.41	8.71	0.94	0.08	0.05	2.03	-0.22	2.62	0.45
G1302	Q6	0.37	0.23	1.89	0.45	1.18	0.21	0.82	0.14	0.92	-0.27
G1302	Q7	3.49	0.66	9.65	1.01	0.70	-0.16	0.32	-0.09	3.49	0.48
G1302	Q8	0.37	0.22	3.09	0.57	1.30	0.21	3.92	0.30	5.03	0.60
G1307	Q1	8.82	0.52	0.50	0.23	18.28	0.29	24.04	0.59	33.15	1.15
G1307	Q2	29.89	0.93	35.21	2.31	33.98	0.39	58.12	0.87	42.88	1.34
G1307	Q3	0.02	0.03	0.79	0.28	88.32	0.65	53.10	0.87	11.17	0.65
G1307	Q4	2.51	0.25	20.71	1.36	38.55	0.41	28.09	0.60	38.83	1.24
G1307	Q5	2.85	0.28	8.71	0.92	91.72	0.68	29.33	0.65	22.69	0.99
G1307	Q6	0.03	0.03	3.06	0.54	30.25	0.38	10.43	0.39	8.23	0.59
G1307	Q7	4.50	0.36	2.79	0.54	2.48	-0.11	0.29	0.06	22.54	0.98
G1307	Q8	24.91	0.85	36.75	2.04	7.63	0.18	5.58	0.27	64.53	1.58
G1402	Q1	1.79	-0.34	1.57	0.40	9.26	0.56	2.18	0.21	5.18	0.70
G1402	Q2	7.40	-0.68	0.04	-0.06	13.31	0.69	7.11	0.39	1.64	0.36
G1402	Q3	1.19	0.30	17.95	1.31	1.44	0.22	2.56	-0.23	6.56	0.73

Reference		CA ABA									
Focal		CA Accredited		CA Registered		Out-of-state ABA		Attorney		Foreign JD	
Year	Item	MH Statistic	LA- LOR	MH Statistic	LA- LOR	MH Statistic	LA- LOR	MH Statistic	LA- LOR	MH Statistic	LA- LOR
G1402	Q4	0.67	-0.22	1.52	0.39	0.77	0.16	2.08	-0.21	0.11	-0.09
G1402	Q5	2.09	0.39	12.50	1.20	10.74	0.60	10.77	0.46	9.66	0.87
G1402	Q6	0.71	0.22	1.19	0.33	16.15	0.79	8.39	0.42	2.23	0.42
G1402	Q7	5.62	0.62	2.25	0.47	2.04	0.25	0.08	0.04	13.71	1.14
G1402	Q8	1.62	0.34	6.96	0.87	0.66	0.15	0.14	-0.05	12.71	0.97
G1407	Q1	4.19	0.36	9.86	0.89	7.69	0.20	19.82	0.52	39.92	1.46
G1407	Q2	0.02	-0.02	30.03	1.71	50.75	0.50	56.46	0.84	31.57	1.21
G1407	Q3	8.99	0.53	8.81	0.88	1.58	0.09	0.33	0.07	8.94	0.65
G1407	Q4	0.64	0.13	17.05	1.15	62.91	0.56	64.97	0.92	35.54	1.31
G1407	Q5	10.04	0.55	4.62	0.62	15.00	0.28	17.71	0.47	15.40	0.84
G1407	Q6	19.46	0.75	2.07	0.41	12.74	0.25	15.44	0.43	50.32	1.54
G1407	Q7	7.83	0.47	15.37	1.13	5.92	0.17	24.56	0.54	44.55	1.37
G1407	Q8	24.94	0.85	16.99	1.08	4.79	0.15	13.53	0.40	36.45	1.28
G1502	Q1	3.82	0.58	5.90	0.76	5.68	0.47	13.09	0.57	0.35	0.14
G1502	Q2	5.40	0.64	15.11	1.21	17.72	0.79	19.91	0.68	3.77	0.45
G1502	Q3	0.92	0.28	0.76	0.27	0.22	-0.09	0.30	-0.09	0.01	0.02
G1502	Q4	1.82	0.38	0.49	-0.19	23.30	1.01	24.00	0.78	7.65	0.70
G1502	Q5	2.56	0.47	0.93	0.30	0.12	0.07	4.30	-0.32	0.59	-0.18
G1502	Q6	1.67	-0.36	0.71	-0.25	4.09	0.41	1.18	0.17	2.24	0.38
G1502	Q7	2.09	0.43	12.87	1.01	0.01	0.02	0.08	0.05	3.73	0.48
G1502	Q8	2.54	0.42	14.87	1.16	3.03	0.33	0.27	-0.08	11.94	0.83
G1507	Q1	2.57	0.34	29.37	2.13	22.41	0.34	33.36	0.68	45.48	1.34
G1507	Q2	1.92	0.28	21.56	1.72	41.33	0.46	21.44	0.53	28.98	1.00
G1507	Q3	9.79	0.66	23.02	1.71	39.77	0.47	84.65	1.22	53.87	1.54
G1507	Q4	0.30	0.13	7.00	1.03	87.29	0.70	64.88	0.96	5.07	0.42
G1507	Q5	0.81	0.19	15.09	1.58	20.09	0.33	12.22	0.43	18.00	0.92
G1507	Q6	0.94	-0.20	0.55	0.25	3.64	0.14	25.92	0.64	20.33	0.99
G1507	Q7	8.86	0.64	17.64	1.51	8.17	0.20	19.79	0.51	27.11	1.00
G1507	Q8	14.94	0.82	15.80	1.29	0.02	-0.01	10.32	0.39	26.44	1.01
G1602	Q1	2.31	-0.50	3.71	0.63	3.46	0.39	3.34	0.31	1.00	0.27
G1602	Q2	3.38	-0.59	0.11	-0.11	0.16	0.09	1.24	0.19	8.16	0.75
G1602	Q3	0.93	-0.32	1.09	0.33	11.80	0.75	22.67	0.84	7.39	0.78
G1602	Q4	0.01	-0.04	3.43	0.64	0.12	0.08	0.42	-0.11	0.47	-0.19
G1602	Q5	0.20	0.17	3.57	0.63	11.77	0.77	4.99	0.40	13.05	1.08
G1602	Q6	0.04	0.07	0.04	0.07	4.34	0.45	1.77	0.23	1.49	0.35
G1602	Q7	1.08	0.40	0.45	0.27	1.10	-0.23	0.90	-0.17	6.27	0.68
G1602	Q8	2.12	0.46	5.31	0.70	4.26	0.43	0.26	0.09	20.45	1.18
G1607	Q1	1.28	0.26	6.30	0.77	3.71	0.16	3.18	0.24	17.36	0.75
G1607	Q2	0.10	0.07	11.31	1.15	4.04	0.16	12.48	0.49	1.80	0.27
G1607	Q3	3.38	0.39	7.39	0.83	18.33	0.35	23.75	0.70	30.47	1.25

Reference		CA ABA									
Focal		CA Accredited		CA Registered		Out-of-state ABA		Attorney		Foreign JD	
Year	Item	MH Statistic	LA- LOR	MH Statistic	LA- LOR	MH Statistic	LA- LOR	MH Statistic	LA- LOR	MH Statistic	LA- LOR
G1607	Q4	0.37	0.14	17.76	1.42	0.20	0.04	10.33	0.44	17.15	0.86
G1607	Q5	0.22	0.11	9.12	0.93	24.97	0.40	23.98	0.70	22.66	0.93
G1607	Q6	11.17	0.80	19.67	1.47	16.51	0.33	21.13	0.62	4.81	0.44
G1607	Q7	13.37	0.98	17.25	1.56	6.56	-0.21	0.00	-0.01	22.90	1.00
G1607	Q8	23.74	1.23	42.64	2.37	0.27	0.04	13.61	0.48	41.30	1.23
G1702	Q1	0.31	0.17	2.05	0.57	13.09	0.89	2.01	0.27	7.96	0.72
G1702	Q2	0.02	-0.05	1.48	0.50	16.20	1.06	4.57	0.42	11.12	0.92
G1702	Q3	3.25	0.64	0.35	0.23	14.12	0.95	22.24	0.94	9.56	0.91
G1702	Q4	7.00	0.96	4.18	0.85	4.41	0.53	0.00	0.01	0.38	0.17
G1702	Q5	0.98	0.33	3.08	0.78	13.99	1.02	12.94	0.73	13.31	1.07
G1702	Q6	0.78	0.30	0.26	0.19	12.46	0.90	13.55	0.74	17.09	1.07
G1702	Q7	1.63	0.42	1.59	0.54	1.12	-0.25	1.10	0.20	0.01	-0.03
G1702	Q8	4.26	0.73	1.53	0.52	6.10	0.60	0.56	0.15	9.84	0.85
G1707	Q1	0.56	-0.13	7.60	1.00	80.97	0.71	76.03	1.00	28.54	0.77
G1707	Q2	18.26	0.77	16.52	1.47	10.11	0.24	11.97	0.38	27.08	0.77
G1707	Q3	2.68	0.28	16.52	1.40	27.35	0.40	97.49	1.11	25.59	0.75
G1707	Q4	1.12	0.18	4.86	0.73	7.30	0.20	10.35	0.34	35.77	0.90
G1707	Q5	2.68	0.28	6.70	0.85	7.99	0.21	50.52	0.78	88.33	1.48
G1707	Q7	21.31	0.79	12.11	1.04	0.07	0.02	29.12	0.57	117.46	1.58
G1802	Q1	5.59	0.72	1.00	0.33	1.53	0.27	0.01	-0.01	8.01	0.65
G1802	Q2	0.60	-0.22	0.77	0.29	1.45	0.27	0.15	0.07	3.81	0.45
G1802	Q3	3.27	0.54	2.72	0.52	0.33	0.12	0.24	-0.09	6.45	0.55
G1802	Q4	4.96	0.65	2.27	0.47	10.68	0.74	8.62	0.54	35.86	1.44
G1802	Q5	1.68	-0.38	8.67	1.01	2.34	0.34	7.15	0.47	1.52	0.27
G1802	Q7	2.83	0.46	5.10	0.79	1.57	0.28	0.07	-0.05	12.51	0.80
G1807	Q1	1.66	0.18	3.73	0.49	23.05	0.35	19.29	0.49	50.74	0.77
G1807	Q2	16.85	0.58	11.97	0.89	73.04	0.61	92.37	1.11	174.53	1.54
G1807	Q3	32.57	0.84	22.18	1.14	47.88	0.49	62.91	0.92	99.09	1.08
G1807	Q4	2.44	0.21	1.64	0.31	112.95	0.76	57.72	0.87	75.03	0.98
G1807	Q5	7.64	0.41	8.86	0.74	30.75	0.40	33.54	0.66	40.90	0.72
G1807	Q7	36.31	0.86	32.51	1.38	18.77	0.30	24.00	0.53	173.05	1.42
G1902	Q1	1.94	0.37	0.12	0.09	3.58	0.40	4.90	0.40	7.80	0.53
G1902	Q2	0.55	0.19	0.02	0.04	0.50	0.15	0.61	-0.14	18.95	0.81
G1902	Q3	0.04	-0.05	0.17	-0.11	3.39	0.38	12.33	0.62	29.50	1.08
G1902	Q4	0.35	0.15	7.05	0.76	4.95	0.49	0.02	0.02	17.49	0.84
G1902	Q5	0.01	0.02	2.51	0.46	11.16	0.72	3.12	0.32	26.64	1.06
G1902	Q7	5.20	0.58	8.38	0.79	1.24	0.23	3.50	-0.33	20.35	0.85

APPENDIX D: RACE/ETHNICITY BY SUBJECT RESULTS

Caucasian-Black/AA and Subject DIF Statistics

Subject	MH Statistic	LA-LOR Statistic					
		Favors Black/AA			Favors Caucasian		
		Large	Moderate	Negligible	Negligible	Moderate	Large
Business	Nonsig.	0	0	2	4	2	0
Associations	Significant	0	0	0	3	0	0
Civil	Nonsig.	0	1	1	8	0	0
Procedure	Significant	0	0	0	0	2	2
Community	Nonsig.	0	0	0	6	1	0
Property	Significant	0	0	0	3	1	0
Constitutional	Nonsig.	0	0	0	8	0	0
Law	Significant	0	0	0	0	2	0
Contracts	Nonsig.	0	0	2	9	4	0
	Significant	0	0	0	4	3	3
Criminal Law	Nonsig.	0	0	1	6	2	0
& Procedure	Significant	0	0	0	2	1	0
Evidence	Nonsig.	0	0	1	4	1	0
	Significant	0	0	0	2	3	0
Immigration	Nonsig.	0	0	0	0	1	0
Law*	Significant	0	0	0	0	0	0
Professional	Nonsig.	0	0	1	10	3	0
Responsibility	Significant	0	0	0	1	1	0
Real Property	Nonsig.	0	0	2	4	2	0
	Significant	0	0	0	0	1	2
Regulatory	Nonsig.	0	0	0	1	0	0
Law*	Significant	0	0	0	1	0	0
Remedies	Nonsig.	0	0	0	2	2	1
	Significant	0	0	0	0	0	1
Torts	Nonsig.	0	0	1	4	2	0
	Significant	0	0	0	4	1	0
Trusts	Nonsig.	0	0	0	2	0	0
	Significant	0	0	0	1	0	0
Wills &	Nonsig.	0	0	0	3	2	0
Succession	Significant	0	0	0	0	0	2

*The Immigration Law and Regulatory Law subject areas were tested via performance test items only and therefore are not listed as subjects tested on the CBX exam.

Caucasian-Asian and Subject DIF Statistics

Subject	MH Statistic	Large	LA-LOR Statistic				
			Favors Asian		Favors Caucasian		
			Moderate	Negligible	Negligible	Moderate	Large
Business	Nonsig.	0	0	7	3	0	0
Associations	Significant	0	0	1	0	0	0
Civil	Nonsig.	0	0	7	3	0	0
Procedure	Significant	0	0	3	1	0	0
Community	Nonsig.	0	0	3	6	0	0
Property	Significant	0	0	0	2	0	0
Constitutional	Nonsig.	0	0	2	7	0	0
Law	Significant	0	0	0	1	0	0
Contracts	Nonsig.	0	0	6	12	0	0
	Significant	0	0	4	1	1	0
Criminal Law	Nonsig.	0	0	2	4	0	0
& Procedure	Significant	0	0	0	5	0	1
Evidence	Nonsig.	0	0	6	3	0	0
	Significant	0	0	0	1	0	1
Immigration	Nonsig.	0	0	0	1	0	0
Law*	Significant	0	0	0	0	0	0
Professional	Nonsig.	0	0	3	6	0	0
Responsibility	Significant	0	0	0	5	1	1
Real Property	Nonsig.	0	0	4	4	0	0
	Significant	0	0	1	0	1	1
Regulatory	Nonsig.	0	0	1	1	0	0
Law*	Significant	0	0	0	0	0	0
Remedies	Nonsig.	0	0	2	3	0	0
	Significant	0	0	0	1	0	0
Torts	Nonsig.	0	0	3	4	0	0
	Significant	0	0	1	1	1	2
Trusts	Nonsig.	0	0	2	1	0	0
	Significant	0	0	0	0	0	0
Wills &	Nonsig.	0	0	2	4	0	0
Succession	Significant	0	0	1	0	0	0

*The Immigration Law and Regulatory Law subject areas were tested via performance test items only and therefore are not listed as subjects tested on the CBX exam.

Caucasian-Hispanic and Subject DIF Statistics

Subject	MH Statistic	LA-LOR Statistic					
		Favors Hispanic			Favors Caucasian		
		Large	Moderate	Negligible	Negligible	Moderate	Large
Business	Nonsig.	0	0	3	8	0	0
Associations	Significant	0	0	0	0	0	0
Civil	Nonsig.	0	0	1	11	0	0
Procedure	Significant	0	0	0	0	0	1
Community	Nonsig.	0	0	8	3	0	0
Property	Significant	0	0	0	0	0	0
Constitutional	Nonsig.	0	0	9	0	0	0
Law	Significant	0	0	0	1	0	0
Contracts	Nonsig.	0	0	6	11	0	0
	Significant	0	0	0	6	1	1
Criminal Law	Nonsig.	0	0	3	6	0	0
& Procedure	Significant	0	0	1	1	0	0
Evidence	Nonsig.	0	0	3	6	0	0
	Significant	0	0	0	1	1	0
Immigration	Nonsig.	0	0	0	1	0	0
Law*	Significant	0	0	0	0	0	0
Professional	Nonsig.	0	0	10	5	0	0
Responsibility	Significant	0	0	0	1	0	0
Real Property	Nonsig.	0	0	3	3	0	0
	Significant	0	0	0	2	2	1
Regulatory	Nonsig.	0	0	0	2	0	0
Law*	Significant	0	0	0	0	0	0
Remedies	Nonsig.	0	0	2	4	0	0
	Significant	0	0	0	0	0	0
Torts	Nonsig.	0	0	5	6	0	0
	Significant	0	0	0	1	0	0
Trusts	Nonsig.	0	0	1	1	0	0
	Significant	0	0	0	1	0	0
Wills &	Nonsig.	0	0	1	6	0	0
Succession	Significant	0	0	0	0	0	0

*The Immigration Law and Regulatory Law subject areas were tested via performance test items only and therefore are not listed as subjects tested on the CBX exam.

Caucasian-Other and Subject DIF Statistics

Subject	MH Statistic	Large	LA-LOR Statistic				
			Favors Other		Favors Caucasian		
			Moderate	Negligible	Negligible	Moderate	Large
Business	Nonsig.	0	0	5	6	0	0
Associations	Significant	0	0	0	0	0	0
Civil	Nonsig.	0	2	6	5	0	1
Procedure	Significant	0	0	0	0	0	0
Community	Nonsig.	0	0	2	6	2	0
Property	Significant	0	0	0	0	1	0
Constitutional	Nonsig.	0	1	2	5	1	0
Law	Significant	0	0	0	0	1	0
Contracts	Nonsig.	0	2	8	10	1	3
	Significant	0	0	0	0	1	0
Criminal Law	Nonsig.	0	0	2	9	0	1
& Procedure	Significant	0	0	0	0	0	0
Evidence	Nonsig.	0	0	2	4	0	4
	Significant	0	0	0	0	1	0
Immigration	Nonsig.	0	0	0	1	0	0
Law*	Significant	0	0	0	0	0	0
Professional	Nonsig.	0	1	4	7	1	1
Responsibility	Significant	0	0	0	0	0	2
Real Property	Nonsig.	0	0	5	3	1	2
	Significant	0	0	0	0	0	0
Regulatory	Nonsig.	1	0	0	1	0	0
Law*	Significant	0	0	0	0	0	0
Remedies	Nonsig.	0	0	2	3	1	0
	Significant	0	0	0	0	0	0
Torts	Nonsig.	0	1	2	7	0	1
	Significant	0	0	0	0	1	0
Trusts	Nonsig.	0	1	0	1	0	1
	Significant	0	0	0	0	0	0
Wills &	Nonsig.	1	0	3	2	1	0
Succession	Significant	0	0	0	0	0	0

*The Immigration Law and Regulatory Law subject areas were tested via performance test items only and therefore are not listed as subjects tested on the CBX exam.

Attachment C

THE STATE BAR OF CALIFORNIA
CALIFORNIA BAR EXAMINATION

CONFIDENTIAL



DIVISION OF PROGRAMS AND POLICY REVIEW

THE STATE BAR OF CALIFORNIA

REVIEW OF THE CALIFORNIA BAR EXAMINATION ADMINISTRATION AND ASSOCIATED COMPONENTS



CONFIDENTIAL

January 2020

This report was prepared and written by the
Division of Programs and Policy Review (DPPR)
and the Office of Professional Examination Services (OPES)
California Department of Consumer Affairs

Tracy A. Montez, Ph.D., Chief, DPPR

Karen Okicich, M.A., Research Data Supervisor II, OPES



EXECUTIVE SUMMARY

The State Bar of California (State Bar) requested that the Department of Consumer Affairs (DCA) Division of Programs and Policy Review (DPPR) complete a review of the administration of the California Bar Examination, which is administered by the State Bar.

Specifically, the DPPR review had three purposes:

1. To evaluate test administration, grader data entry and data processing protocols, communication with candidates and other stakeholders, and test security associated with the California Bar Examination, as well as contingency planning for exam administration.
2. To conduct a gap analysis comparing the current process for administration of the California Bar Examination with best practices outlined in the *Standards for Educational and Psychological Testing (2014) (Standards)*.¹
3. To provide recommendations for the modification of the current process of administration of the California Bar Examination for conforming that process to the practices specified in the *Standards*.

State Bar representatives and the DPPR Chief, Dr. Montez, participated in a project kick-off conference call on August 29, 2019 to review the scope of the project and associated expectations.

It is important to note that the purpose of this evaluation did not include an analysis of the underlying components to support the validity of the California Bar Examination (i.e., occupational analysis (OA),² examination development, passing scores,³ and examination performance). However, relevant standards are included to facilitate additional reviews, as needed.

The State Bar should be commended for their willingness to release information about processes and procedures for constructive evaluation. It is obvious that State Bar management and staff are committed to building and administering a fair, reliable, and valid examination program.

¹ *Standards* references information taken from: American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. *Standards for Educational and Psychological Testing (2014)*. Washington, DC: American Educational Research Association.

² An occupational analysis is also known as a job analysis, practice analysis, or task analysis.

³ A passing score is also known as a pass point or cut score.

Suggested recommendations are listed in the final chapter and are interrelated within the cycle of examination validation (See Figure 1).

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION.....	1
CHAPTER 2 OCCUPATIONAL ANALYSIS	3
CHAPTER 3 EXAMINATION DEVELOPMENT	6
CHAPTER 4 PASSING SCORES AND PASSING RATES.....	8
CHAPTER 5 TEST ADMINISTRATION	12
CHAPTER 6 GRADER TRAINING, SCORING, AND PERFORMANCE STANDARDS	18
CHAPTER 7 TEST SECURITY.....	25
CHAPTER 8 COMMUNICATION TO CANDIDATES AND STAKEHOLDERS.....	28
CHAPTER 9 CONCLUSIONS.....	29
CHAPTER 10 REFERENCES.....	32

TABLE OF FIGURES

FIGURE 1. EXAMINATION DEVELOPMENT CYCLE	5
FIGURE 2. LEVELS OF COGNITIVE PROCESSING.....	7
FIGURE 3. NORM-REFERENCED PASSING STANDARD	9
FIGURE 4. CRITERION-REFERENCED PASSING STANDARD.....	10

CHAPTER 1 | INTRODUCTION

PURPOSE OF THE COMPREHENSIVE REVIEW

The State Bar of California (State Bar) requested that the Department of Consumer Affairs (DCA) Division of Programs and Policy Review (DPPR) complete a review of the administration of the California Bar Examination, which is administered by the State Bar.

The State Bar of California's mission is to protect the public and includes the primary functions of licensing, regulation and discipline of attorneys; the advancement of the ethical and competent practice of law; and support of efforts for greater access to, and inclusion in, the legal system. (<http://www.calbar.ca.gov/About-Us/Our-Mission>)

The State Bar licenses attorneys and regulates the profession and practice of law in California. Practicing lawyers must pass the California Bar Examination to be licensed by the State Bar. As of July 2017, the format of the California Bar Examination⁴ is constructed as follows:

General Bar Examination

The examination will be administered over two days with the following components: one morning session during which three one-hour essay questions will be administered; one afternoon session during which two one-hour essay questions and one 90-minute Performance Test will be administered; and morning and afternoon sessions consisting of three hours each, during which 100 multiple-choice items for each session will be administered (the Multistate Bar Examination (MBE)).

Attorneys' Examination

The examination will be administered over one day with the following components: one morning session during which three one-hour essay questions will be administered; and one afternoon session during which two one-hour essay questions and one 90-minute Performance Test will be administered.

The DPPR review had three purposes:

1. To evaluate test administration, grader data entry and data processing protocols, communication with candidates and other stakeholders, and test security associated with the California Bar Examination, as well as contingency planning for exam administration.

⁴ <http://www.calbar.ca.gov/Admissions/Examinations/California-Bar-Examination/Changes-to-Bar-Exam>

2. To conduct a gap analysis comparing the current process for administration of the California Bar Examination with best practices outlined in the *Standards for Educational and Psychological Testing (2014) (Standards)*.⁵
3. To provide recommendations for the modification of the current process of administration of the California Bar Examination for conforming that process to the practices specified in the *Standards*.

State Bar representatives and the DPPR Chief, Dr. Montez, participated in a project kick-off conference call on August 29, 2019 to review the scope of the project and associated expectations. Expectations included, for example, how to address subversion incidents during administration of the California Bar Examination and the impact on security of extending testing over multiple days as an accommodation.

During the conference call, DPPR requested documentation from the State Bar to determine whether the following California Bar Examination components met professional guidelines and technical standards outlined in the *Standards*: (a) test administration, (b) grader data entry and data processing protocols, (c) test security, and (d) communication with candidates and stakeholders. This documentation was to be provided within a secure shared drive created by the State Bar.

It is important to note that the purpose of this evaluation did not include an analysis of the underlying components to support the validity of the California Bar Examination (i.e., occupational analysis (OA),⁶ examination development, passing scores,⁷ and examination performance). However, relevant standards are included to facilitate additional reviews, as needed.

FORMAT OF THE REPORT

The chapters of this report provide the relevant standards related to the evaluation purpose and describe the findings and recommendations that DPPR identified during its review. Note that the report documents both *findings*, important factors that are compliant with technical standards and guidelines, and *recommendations*, suggestions that should be considered to maintain compliance with professional guidelines and technical standards.

⁵ *Standards* references information taken from: American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. *Standards for Educational and Psychological Testing (2014)*. Washington, DC: American Educational Research Association.

⁶ An occupational analysis is also known as a job analysis, practice analysis, or task analysis.

⁷ A passing score is also known as a pass point or cut score.

CHAPTER 2 | OCCUPATIONAL ANALYSIS

STANDARDS

The following standard is most relevant to conducting OAs for licensing examinations, as referenced in the *Standards*.

Standard 11.13

The content domain to be covered by a credentialing test should be defined clearly and justified in terms of the importance of the content for credential-worthy performance in an occupation or profession. A rationale and evidence should be provided to support the claim that the knowledge or skills being assessed are required for credential-worthy performance in that occupation and are consistent with the purpose for which the credentialing program was instituted (pp. 181-182).

The comment following Standard 11.13 emphasizes its relevance:

Comment: Typically, some form of job or practice analysis provides the primary basis for defining the content domain. If the same examination is used in the credentialing of people employed in a variety of settings and specialties, a number of different job settings may need to be analyzed. Although the job analysis techniques may be similar to those used in employment testing, the emphasis for credentialing is limited appropriately to knowledge and skills necessary for effective practice. . . .

In tests used for licensure, knowledge and skills that may be important to success but are not directly related to the purpose of licensure (e.g., protecting the public) should not be included (p. 182).

FINDINGS

In December 2018, the State Bar began the first California-specific study of the knowledge and skills needed by entry-level attorneys. To date, the study has collected detailed, empirical data about how attorneys use their knowledge and skills to perform tasks in their legal practices. Key components of the study include focus groups and surveys of California attorneys. Scantron⁸ is conducting the study. A working group, with members selected by the California Supreme Court from state and national stakeholder groups, oversees the study.

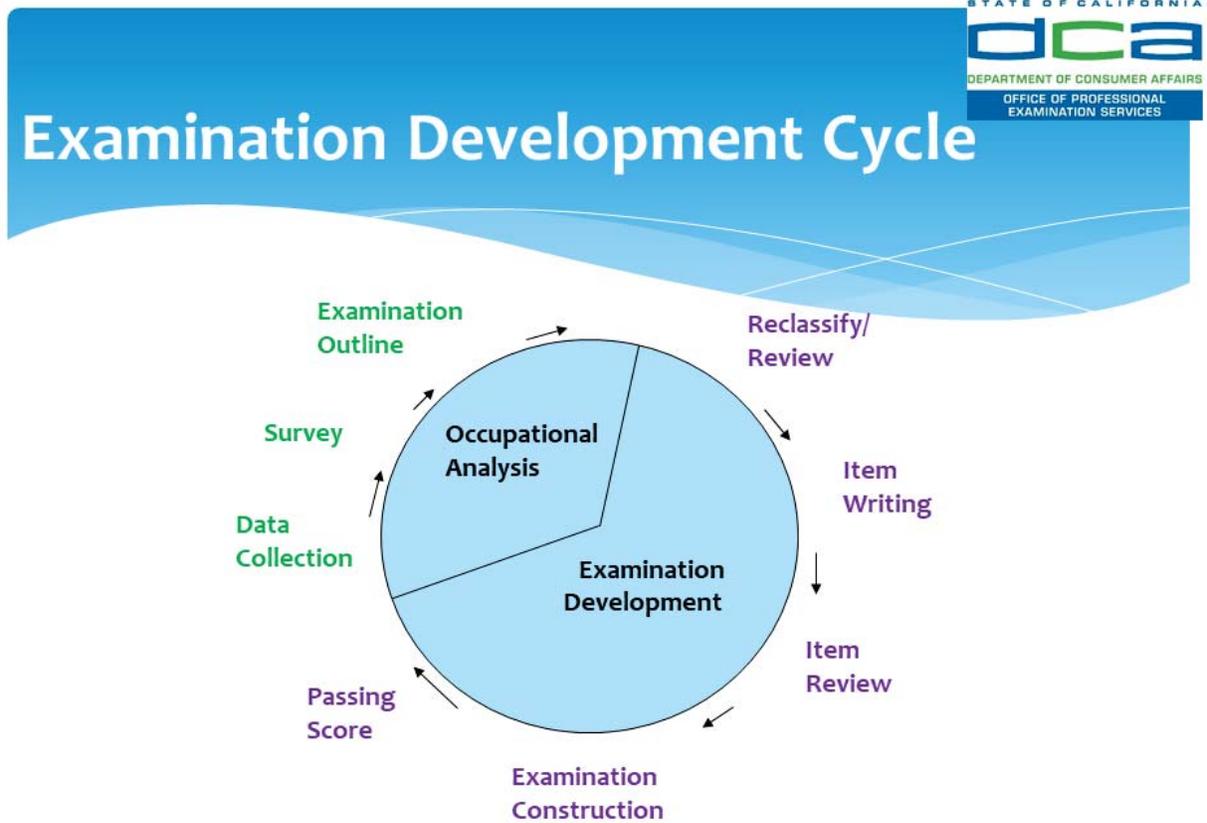
The final report, scheduled for completion by December 2019, will set the foundation for revisiting the California Bar Examination passing standards, content, format, and other aspects

⁸ Scantron acquired Castle Worldwide in June 2018.

of the test (<https://www.calbar.ca.gov/Admissions/Examinations/California-Bar-Examination/California-Bar-Examination-Studies>).

For informational purposes, Figure 1 shows that an OA is a critical component of the examination development or validation cycle. It is important to note that the cycle represents ongoing efforts to maintain the fairness, validity, and legal defensibility of an examination. According to the *Uniform Guidelines on Employee Selection Procedures* (1978), the time interval to conduct a job analysis for updating knowledge and skills varies, but there will seldom be more than five years between analyses. For DCA, section 139 (c) and (d) of the Business and Professions (B&P) Code mandates that DCA annually report to the California Legislature on the methods used by each regulatory entity for ensuring that its licensing examinations are periodically evaluated. Further, B&P Code section 139 (a) and (b) directed DCA to develop a policy regarding examination development and validation. The Licensure Examination Validation Policy (OPES 18-02) states that OAs are fundamental components of a licensure program. OPES 18-02 further states that licensure examinations with substantial validity evidence are essential in preventing unqualified individuals from obtaining professional licenses. To that end, licensure examinations must be: (a) developed following an examination outline that is based on a current OA; (b) regularly evaluated; and (c) updated when tasks performed or prerequisite knowledge in a profession or on a job change (i.e., about every five years), or to prevent overexposure of test questions.

FIGURE 1. EXAMINATION DEVELOPMENT CYCLE



CHAPTER 3 | EXAMINATION DEVELOPMENT

STANDARDS

Examination development includes many steps within an examination program, from the development of an examination outline to scoring and analyzing items after the administration of an examination.

The following standards are most relevant to examination development for licensing examinations, as referenced in the *Standards*.

Standard 4.7

The procedures used to develop, review, and try out items and to select items from the item pool should be documented (p. 87).

Standard 4.12

Test developers should document the extent to which the content domain of a test represents the domain defined in the test specifications (p. 89).

FINDINGS

Although examination development was not a component of the project, recommendations will be offered. These recommendations are suggested in response to general information shared during phone interviews and review of documents provided by the State Bar.

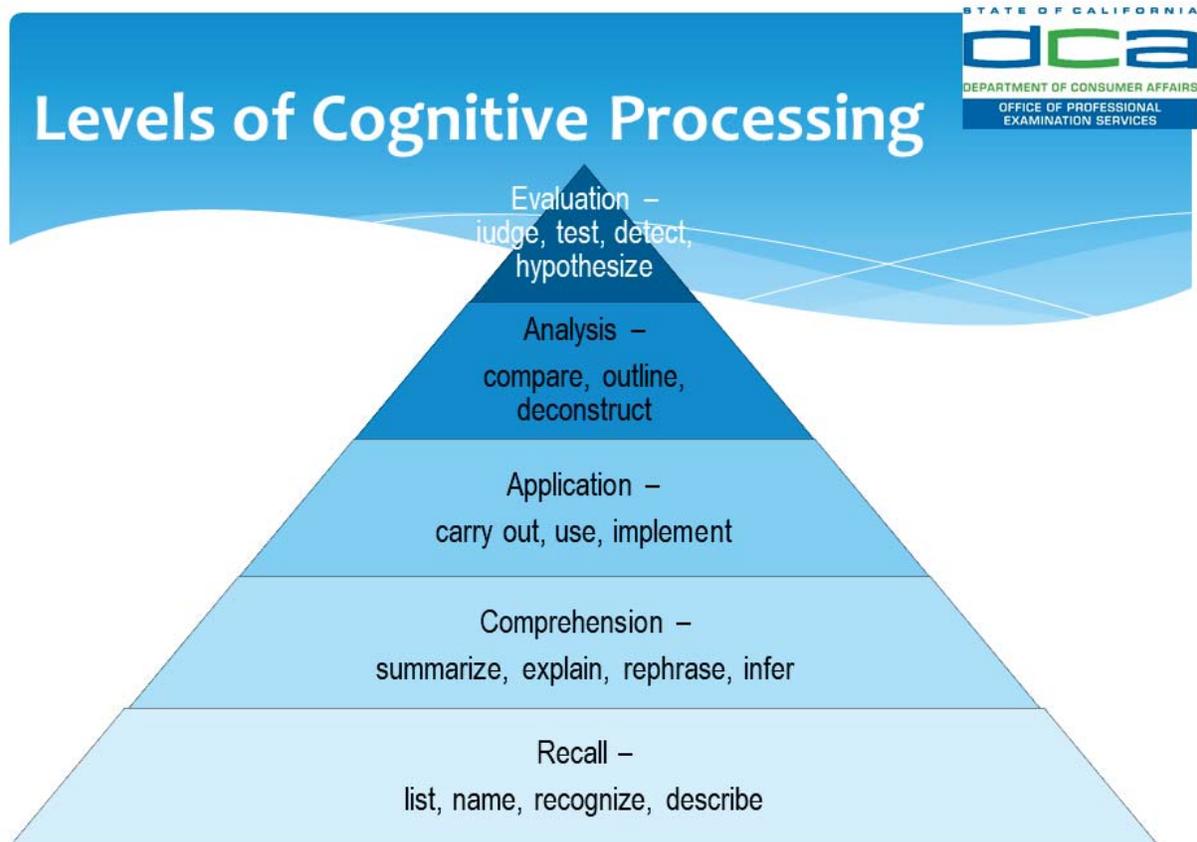
RECOMMENDATIONS

These recommendations are presented below and discussed in greater detail in Chapter 9 – Conclusions.

- a. The examination outline or test specification resulting from the practice analysis should stimulate a discussion about the most fair, reliable, and valid tool to assess entry-level competence to practice as an attorney in California.
- b. Then, a systematic plan for examination development should be created. The plan should be based on the current practice analysis, be ongoing, use subject matter experts, and focus on minimum acceptable competence standards (i.e., entry-level).
- c. The plan should also include the time frame for administering the assessment tool(s). For example, the plan should state how often to offer the assessment (e.g., quarterly, twice a year, etc.) and the time frame for administering the component(s) of the assessment (e.g., 60-minute multiple choice test composed of 50 and 60 minutes per essay question).

- d. A policy on examination validation should be formulated to guide future examination development activities, including scoring and administration. This policy should be based on professional testing guidelines and technical standards. As an example of a guideline, Figure 2 shows one variation of the levels of cognitive processing. For licensing examinations, questions should be written to require higher levels of cognitive processing (i.e., thinking) when answering or responding to questions. Candidates should be challenged to apply, analyze, and evaluate information. Merely recognizing answers or explaining in general terms does not adequately assess entry-level skill for ensuring safe and competent practice in a given profession. The policy should address questions appropriate for use on licensure examinations versus academic examinations.

FIGURE 2. LEVELS OF COGNITIVE PROCESSING



CHAPTER 4 | PASSING SCORES AND PASSING RATES

STANDARDS

The passing score of an examination is the score that represents the level of performance that divides those candidates for licensure who are minimally competent from those who are not competent.

The following standards are most relevant to passing scores, cut points, or cut scores for licensing examinations, as referenced in the *Standards*.

Standard 5.21

When proposed score interpretations involve one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly (p. 107).

Standard 11.16

The level of performance required for passing a credentialing test should depend on the knowledge and skills necessary for credential-worthy performance in the occupation or profession and should not be adjusted to control the number or proportion of persons passing the test (p. 182).

The supporting commentary on passing or cut scores for Chapter 5 of the *Standards*, “Scores, Scales, Norms, Score Linking, and Cut Scores,” states that the standard-setting process used should be clearly documented and defensible. The qualifications of the judges [subject matter experts; SMEs] involved and the process of selecting them should be part of the documentation. A sufficiently large and representative group of judges [SMEs] should be involved, and care must be taken to ensure that judges [SMEs] understand the process and procedures they are to follow (p. 101).

In addition, the supporting commentary for Chapter 11 of the *Standards*, “Workplace Testing and Credentialing,” states that the focus of tests used in credentialing is on “the standards of competence needed for effective performance (e.g., in licensure this refers to safe and effective performance in practice)” (p. 175). Further, it states, “Standards must be high enough to ensure that the public, employers, and government agencies are well served, but not so high as to be unreasonably limiting” (p. 176).

FINDINGS

An extensive explanation of the grading process, including the difference between essay and performance questions, and scaling, is provided on the State Bar website at:

<http://www.calbar.ca.gov/Admissions/Examinations/California-Bar-Exam/Description-and-Grading-of-the-California-Bar-Exam>.

For informational purposes, Figures 3 and 4 are presented to distinguish between norm-referenced and criterion-referenced standards for scoring. According to professional guidelines and technical standards, criterion-referenced standards should be used to establish passing scores for licensure examinations. Since most individuals are familiar with norm-referenced standards (e.g., grading on the curve), it is critical for test developers to thoroughly explain methods used to establish criterion-referenced passing scores. See Chapter 6 – Grader Training, Scoring, and Performance Standards for additional explanation.

FIGURE 3. NORM-REFERENCED PASSING STANDARD

Norm-Referenced Passing Standard



- Measures performance based on the candidate distribution
- Allows for comparison among test takers
 - Example: Grading on the curve (highest score becomes maximum achievable)
- Some candidates will pass regardless of true ability (competence)

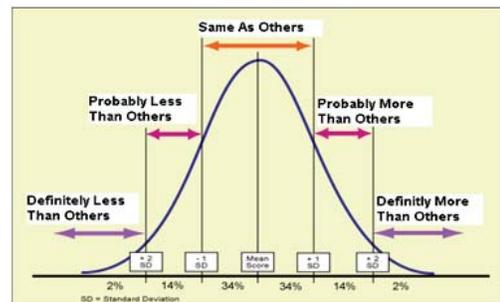
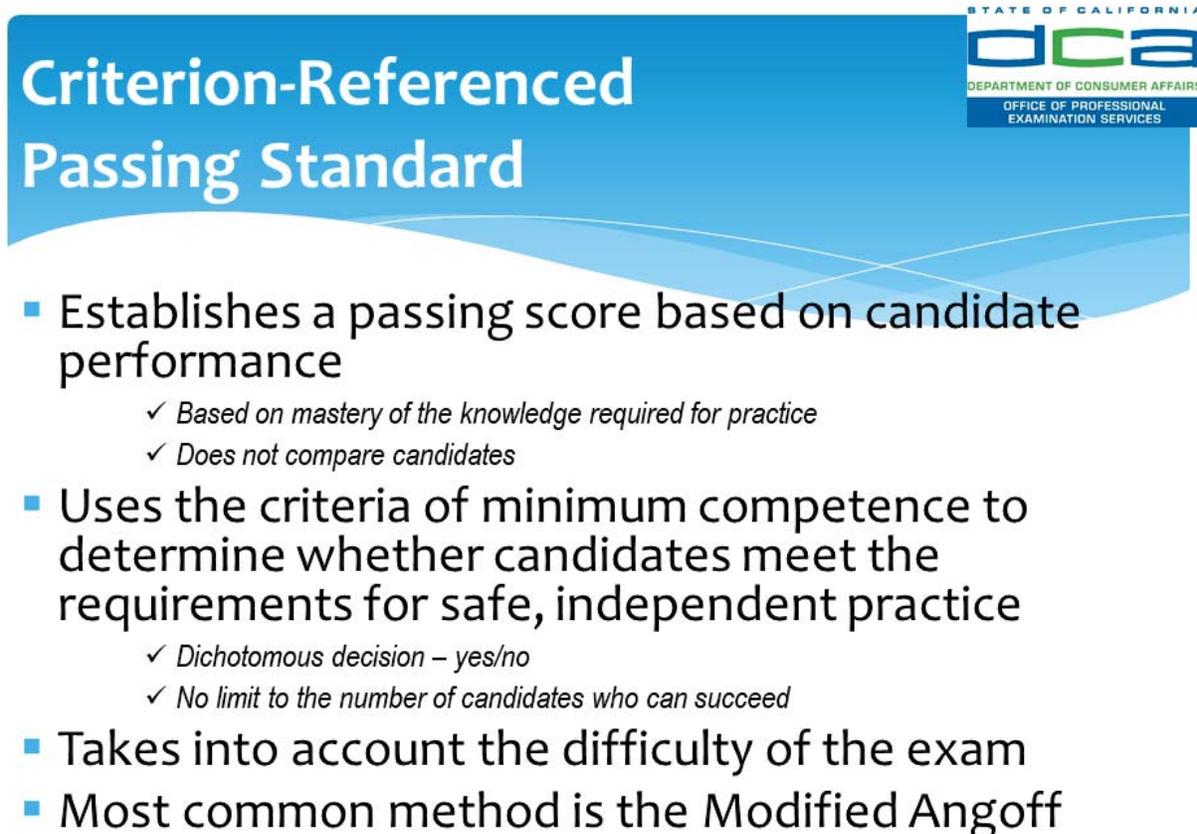


FIGURE 4. CRITERION-REFERENCED PASSING STANDARD



Criterion-Referenced Passing Standard

STATE OF CALIFORNIA
dca
DEPARTMENT OF CONSUMER AFFAIRS
OFFICE OF PROFESSIONAL EXAMINATION SERVICES

- Establishes a passing score based on candidate performance
 - ✓ *Based on mastery of the knowledge required for practice*
 - ✓ *Does not compare candidates*
- Uses the criteria of minimum competence to determine whether candidates meet the requirements for safe, independent practice
 - ✓ *Dichotomous decision – yes/no*
 - ✓ *No limit to the number of candidates who can succeed*
- Takes into account the difficulty of the exam
- Most common method is the Modified Angoff

“The defensibility of and the strength of the validity evidence for passing scores relies on the reasonableness of the unbiased process, its rationale and research basis, and the psychometric characteristics of expert judgement. (Downing, p. 20).

The California Bar Examination essay and Performance Test questions are developed independently by members of the Edge Team who then submit the questions to the Committee of Bar Examiners for consideration. Prior to the first calibration meeting, the question developer’s analysis is provided to the graders for helping determine what issues are important in grading the questions.

Since there is no historical information documenting entry-level performance standards, it appears standards fluctuate each time the essay and Performance Test are administered and scored. There needs to be documentation to standardize the development and grading of the essay and Performance Test questions. Further, recently licensed attorneys should be involved in examination development and grading phases.

The grading process described by State Bar is defensible and is closer to a criterion-referenced methodology, but additional steps need to be taken to ensure entry-level standards are documented, used in both examination development and grading.

CHAPTER 5 | TEST ADMINISTRATION

STANDARDS

The following standards are most relevant to standardizing the test administration process for licensing examinations, as referenced in the *Standards*.

Standard 3.4

Test takers should receive comparable treatment during the test administration and scoring process (p. 65).

Standard 4.15

The directions for test administration should be presented with sufficient clarity so that it is possible for others to replicate the administration conditions under which the data on reliability, validity, and (where appropriate) norms were obtained. Allowable variations in administration procedures should be clearly described. The process for reviewing requests for additional testing variations should also be documented (p. 90).

Standard 4.16

The instructions presented to test takers should contain sufficient detail so that test takers can respond to a task in the manner that the test developer intended. When appropriate, sample materials, practice or sample questions, criteria for scoring, and a representative item identified with each item format or major area in the test's classification or domain should be provided to the test takers prior to the administration of the test or should be included in the testing material as part of the standard administration instructions (p. 90).

Standard 6.1

Test administrators should follow carefully the standardized procedures for administration and scoring specified by the test developer and any instructions from the test user (p. 114).

Standard 6.2

When formal procedures have been established for requesting and receiving accommodations, test takers should be informed of these procedures in advance of testing (p. 115).

Standard 6.3

Changes or disruptions to standardized test administration procedures or scoring should be documented and reported to the test user (p. 115).

Standard 6.4

The testing environment should furnish reasonable comfort with minimal distractions to avoid construct-irrelevant variance (p. 116).

Standard 6.5

Test takers should be provided appropriate instructions, practice, and other support necessary to reduce construct-irrelevant variance (p. 116).

Standard 8.1

Information about test content and purposes that is available to any test taker prior to testing should be available to all test takers. Shared information should be available free of charge and in accessible formats (p. 133).

Standard 8.2

Test takers should be provided in advance with as much information about the test, the testing process, the intended test use, test scoring criteria, testing policy, availability of accommodations, and confidentiality protection as is consistent with obtaining valid responses and making appropriate interpretations of test scores (p. 134).

FINDINGS

Test Administration – Candidate Registration

The State Bar website includes links to access important test administration information such as the online Admissions Information Management System (AIMS). Candidates register to take the California Bar Examination using AIMS but also have the option of using a paper application. Other informational links include:

- Creating a new registration
- Learning about the California Bar Examination (e.g., test preparation, FAQs)
- Taking the attorney oath
- Listing of important dates and deadlines
- Contacting the Office of Admissions

Finding: The State Bar registration process appears straightforward. The information available to candidates is detailed and thorough. The candidate registration process appears to meet professional guidelines and technical standards.

Test Administration – Accommodation Requests

Candidates can request accommodations via the online testing accommodations application. The candidate is instructed to discuss with their physician or specialist what accommodation is necessary to allow them to compete on an equal basis with all other applicants. A checklist is

provided to facilitate the request process (see <http://www.calbar.ca.gov/Portals/0/documents/admissions/TA-Petition-Checklist.pdf>). It should be noted that there is an appeal process and an emergency petition pathway.

It has been reported that requests for testing accommodations are increasing. This trend is consistent with other regulatory licensing examinations. It is important to note accommodations that fundamentally alter the measurement of the skills or knowledge the examination is intended to test should not be provided. When declining an accommodation, the reason why the accommodation will alter measurement of the skill and/or knowledge should be provided in the event there is a different accommodation available that does not alter measurement. It is important to allow accommodations for assuring that tests measure the actual skill and/or knowledge of an individual to perform essential job duties rather than reflect limitations caused by the disability.

Finding: The State Bar testing accommodations process is defensible and appears to meet professional guidelines and technical standards.

Test Administration – Test Centers

Candidates take the California Bar Examination at a contracted test center. These sites vary among approximately 15, and factors such as number of applicants and proctor and site availability determine how many sites will be used. For the February 2020 administration, 6 sites are available in southern California and 4 sites in northern California. Admittance restrictions are placed on some sites, such as a designation as a testing accommodations site only.

Finding: Candidates have access to testing sites in major cities, but travel may be required for a majority of candidates, resulting in additional costs and inconvenience.

Test Administration – Preparation and Distribution of Materials

Once the application period closes, examinations are requisitioned. The MBE is ordered through the National Conference of Bar Examiners (NCBE). The essay and Performance Test exams are printed, using a contracted vendor, and shipped to a State Bar office in Los Angeles for test center distribution. Test materials are examined prior to administration for defects. Test materials are also stored in a secure manner. To date, no known security issues have been reported.

Finding: The State Bar takes reasonable measures to ensure the secure preparation, delivery, and storage of test materials. However, the established administration process is more vulnerable to security breaches and general logistical problems than multiple-choice/vignette formats using a computer-based test administration. Those breaches and problems center around the “chain of security” (i.e., printing, delivery, distribution, and collection of test materials). In fairness to paper-and-pencil testing, computer-based testing has become susceptible to security breaches as technology has advanced. However, the “chain of security” tends to be more vulnerable to security breaches and logistical problems than electronic transmission of test questions and data.

Test Administration – General Instructions to Candidates

The State Bar website provides information about the California Bar Examination. The links provide very detailed information to candidates regarding:

- Format of the California Bar Examination
- Using a laptop computer (e.g., software to download)
- What test takers can and cannot bring to the examination
- How to apply to wear religious headwear during the examination
- Start times/late arrivals/exam attendance policy
- Sanctions

Finding: The general instructions provided to candidates appear straightforward. The information available to candidates is detailed and thorough. However, given the three assessment components (essay questions, Performance Test, multiple-choice questions) and differing instructions, construct-irrelevant variance may result (e.g., increased test anxiety) from the intent to be helpful and transparent.

Test Administration – Use of Laptop Computers

Candidates have the option of handwriting answers or using a personal or rented laptop computer. Most candidates choose to use a laptop. Candidates pay \$152 to use a personal laptop. The State Bar provides detailed information about the use of personal or rented laptop computers located at: <https://www.calbar.ca.gov/Admissions/Examinations/California-Bar-Examination/Using-Laptops>.

Topics include:

- Laptop certification – Exemplify software must be downloaded prior to exam day to prevent internet access during examination.
- Exam information – ExamSoft’s FlexSite internet-based examination delivery option will be used for responses to the essay and Performance Test portions.
- Exam day – In the event of a problem such as a software or hardware malfunction, power failure or interruption, candidates are required to handwrite examination answers.
- ExamSoft information – The certification process is completed based on instructions provided by Exemplify/ExamSoft and includes taking the mock examination.
- Preparing for exam day – Candidates must bring a power cord and a laptop with a charged battery and know the laptop/software for taking the essay and Performance Test, including for downloading responses.
- Checklist – To ensure proper certification of laptop computers for use during the examination, a checklist of the steps that must be taken is provided.

It is important to note that Exemplify software is not compatible with certain testing accommodation software such as Job Access With Speech (JAWS) screen reader and Dragon

voice recognition. Additional steps are taken by staff to check laptops using these types of software. These candidates are also required to bring their own printer.

Finding: The procedures established for use of laptop computers appear to meet professional guidelines and technical standards. To restate, although the information is detailed and thorough, the amount of logistical preparation prior to actually sitting for the examination may introduce construct-irrelevant variance.

Test Administration – Proctors and Standardized Procedures

Proctors are recruited and screened for employment by State Bar staff. Information, including qualifications and requirements, can be found at: <http://www.calbar.ca.gov/About-Us/Careers/San-Francisco-Careers/Examinations-Technician/Exam-Proctors>. Proctors cannot be law clerks, law students, paralegals, or attorneys in any jurisdiction.

The State Bar has developed various manuals (e.g., *Proctor Guidelines*, *Staff Member in Charge and Staff Representative Manual*) to facilitate the standardized administration of the California Bar Examination across sites. In response to feedback from proctors, however, critical tasks lists have been created to summarize the main points of the manuals for easier reference. In addition to the manuals, roles (e.g., Head Proctor, Security Proctor, Staff Member in Charge) and responsibilities (e.g., laptop writers, test accommodations) have been identified for proctor assignment. The goal is to create critical tasks lists for all manuals/roles.

Proctor training across test centers is mostly consistent (i.e., as stated, not observed). Depending on the assignment, training may occur the day before or the morning of the examination.

Finding: The procedures established for the test administration process and testing environment appear to meet professional guidelines and technical standards. State Bar staff clearly recognize the need for standardized procedures to facilitate fairness and reliability of assessment across test centers. Staff, however, are open and responsive to feedback for improving the process.

Finding: The proctor pool is reportedly decreasing, possibly due to low pay, complexity of the recruitment and reimbursement process, and general availability of interested parties. This trend is consistent with regulatory examinations that utilize essay, performance, and practical examination formats.

RECOMMENDATIONS

Recommendation: Consider removing the MBE from the two-day administration and utilizing computer-based testing (CBT). This strategy would likely reduce the cognitive load on candidates and provide greater convenience for scheduling (e.g., local test centers, self-scheduling). This recommendation is further discussed in Chapter 9 – Conclusions.

CONCLUSIONS

Given the findings, the test administration protocols put in place by the State Bar appear to meet professional guidelines and technical standards.

CHAPTER 6 | GRADER TRAINING, SCORING, AND PERFORMANCE STANDARDS

STANDARDS

The following standards are most relevant to grader training, scoring, and performance for licensing examinations, as referenced in the *Standards*.

Standard 2.3

For each total score, subscore, or combination of scores that is to be interpreted, estimates of relevant indices of reliability/precision should be reported (p. 43).

Standard 4.10

When a test developer evaluates the psychometric properties of items, the model used for that purpose (e.g., classical test theory, item response theory, or another model) should be documented. The sample used for estimating item properties should be described and should be of adequate size and diversity for the procedure. The process by which items are screened and the data used for screening, such as item difficulty, item discrimination, or differential item functioning (DIF) for major examinee groups, should also be documented. When model-based methods (e.g., IRT) are used to estimate item parameters in test development, the item response model, estimation procedures, and evidence of model fit should be documented (pp. 88-89).

Standard 4.20

The process for selecting, training, qualifying, and monitoring scorers should be specified by the test developer. The training materials, such as the scoring rubrics and examples of test takers' responses that illustrate the levels on the rubric score scale, and the procedures for training scorers should result in a degree of accuracy and agreement among scorers that allows the scores to be interpreted as originally intended by the test developer. Specifications should also describe processes for assessing scorer consistency and potential drift over time in raters' scoring (p. 92).

Standard 4.21

When test users are responsible for scoring and scoring requires scorer judgment, the test user is responsible for providing adequate training and instruction to the scorers and for examining scorer agreement and accuracy. The test developer should document the expected level of scorer agreement and accuracy and should provide as much technical guidance as possible to aid test users in satisfying this standard (p. 92).

Standard 4.23

When a test score is derived from the differential weighting of items or subscores, the test developer should document the rationale and process used to develop, review, and assign item weights... (p. 93).

Standard 6.8

Those responsible for test scoring should establish scoring protocols. Test scoring that involves human judgment should include rubrics, procedures, and criteria for scoring. When scoring of complex responses is done by computer, the accuracy of the algorithm and processes should be documented (p. 118).

Standard 6.9

Those responsible for test scoring should establish and document quality control processes and criteria. Adequate training should be provided. The quality of scoring should be monitored and documented. Any systematic source of scoring errors should be documented and corrected (p. 118).

FINDINGS

Graders – Selection and Training

The State Bar Committee of Bar Examiners (Committee) maintains a pool of approximately 150 licensed attorneys who serve as potential graders. These attorneys represent the diverse areas of practice in California and are located in the San Francisco Bay Area. The attorneys included in this pool have a minimum of 3 years of experience scoring the written portions of the California Bar Examination and many have over 10 years of experience. The Committee also uses Apprentice Graders, who undergo training and calibration processes. Apprentice Graders are used if a grader is unable to complete their assignments (Committee of Bar Examiners, Description and Grading of the California Bar Examination).

For each grading cycle, six panels of 12 graders are convened to evaluate candidate responses on the five essays and the Performance Test question that comprise the written portion of the California Bar Examination. A member of the Examination Development and Grading Team (EDG Team), a group of former graders, and a member of the Committee oversee these panels (Committee of Bar Examiners, Description and Grading of the California Bar Examination).

Graders are required to participate in three calibration meetings during each examination grading cycle. During the first calibration meeting, graders develop grading guidelines and assign weights to essay components. Graders also review and discuss 15 candidate responses on essay questions from the current exam administration. Graders first work as a group to review and assign a grade for each of the selected responses and reach consensus on a final grade. The graders then independently grade an additional 25 candidate responses. These analyses are calculated as a means of evaluating each group's level of calibration. The results

of these analyses are reviewed at the second calibration meeting the following week (Grading the California Bar Examination Grader Instruction II. First Calibration Meeting).

At the second calibration meeting, held the following week, graders review and finalize grading guidelines and the weights assigned to the essay during the first calibration meeting. Graders also review the results of calibration analyses conducted on the 25 responses graded during the first calibration meeting and resolve discrepancies through discussion (Grading the California Bar Examination Grader Instruction III. Second Calibration Meeting). Panelists then read, grade, and discuss 15 additional candidate responses as a group for additional calibration, and assign final consensus grades. Graders are then provided with grading assignments. The number of essay questions each grader reviews varies based on the number of candidates who sat for the examination; however, estimates provided by the Committee are that each grader will review 140-170 essays a week.

A third calibration session is held mid-grading cycle, during which graders read and grade an additional 15 essay responses. The purpose of this calibration session is to review grading standards and ensure graders are applying the same standard (Grading the California Bar Examination Grader Instruction V. Third Calibration Meeting).

Finding: Graders are selected to represent the diverse areas of legal practice in California. Graders appear to receive adequate training on grading procedures and the application of grading standards although the extent to which minimum competence standards are discussed is not clear.

RECOMMENDATION

Recommendation: The pool of potential graders should include licensed attorneys who are currently practicing in the profession. Educators or those involved in the educational process should not be part of the grading process, even as observers. Those associated with professional associations or other vested interests should also be excluded from grading processes to avoid unintentional bias or influence. Since the purpose of licensure is to assess minimum competence for entry into the profession, graders should include those who represent this target population (i.e., entry-level practitioners licensed less than five years).

It appears that the Committee maintains a large pool of potential graders; however, a sliding-scale monetary incentive is offered to graders based on the number of times they have served in the past. It is recommended that the Committee consider rotation of graders and the inclusion of new graders into the process to ensure diverse perspectives are maintained throughout the process. Further, it appears that all potential graders are located in the Bay Area. Professional guidelines recommend that subject matter experts who participate in the development and scoring of licensing examinations be: (1) currently licensed and actively practicing in the field; (2) in good standing; (3) representative of the diversity of practice or specialty areas involved in the profession;

and (4) representative of practice in California in terms of geographic location and years licensed, with an emphasis on entry level.

Scoring – California Bar Examination

Following administration of the California Bar Examination, graders write and submit an analysis of the essay question or the Performance Test to which they are assigned, as well as an outline summary (Grading the Calibration Bar Examination Grader Instruction I. Analysis and Outline for the Essay Question). Copies of the analysis and outline are sent to the Supervising EDG Team Member and all co-graders. At the first calibration session, graders review the analyses prepared and reach consensus on the issues that should be discussed by candidates in answering the essay questions (Grading the California Bar Examination Grader Instruction II. First Calibration Meeting). In addition, graders reach consensus on the weights assigned to each issue. Applying the standards agreed upon, graders assign raw scores to essay and Performance Test response in 5-point increments, using a scale of 40 to 100. Graders are instructed to base their grading on the content of candidate responses and not to consider quality of handwriting or accuracy of spelling and grammar in assigning a grade. A score of 100 (maximum points) is to be assigned “when the grader believes the applicant has done as well as can reasonably be expected of any applicant on that question” (Committee of Bar Examiners’ Policy Regarding Assignment of Grades to Written Assignments).

Candidate scores on the written portion of the California Bar Examination are combined with scores on the MBE to produce a total score (Description and Grading of the California Bar Examination – General Bar Examination and Attorneys’ Examination). In determining a total score, the Committee uses a scaling procedure to convert scores on the written portion of the California Bar Examination to the same scale of measurement used on the MBE. The Committee’s intended purpose for scaling scores on the written examination is to account for average differences in difficulty and grader performance across different exam administrations. MBE scores are reported on a scale ranging from 0 to 2000 points. A total of 700 raw points is possible for the written portion of the exam: 100 points for each of the five essay questions and 200 points for the Performance Test, which are then translated to the same 2000-point scale used for the MBE. A candidate’s total score is the scaled MBE score (on the 2000-point scale) multiplied by .50 plus the converted score on the written section multiplied by .50. The effect of scaling is that the MBE and written sections contribute equal weight to the total score candidates receive (Written=50%, MBE=50%).

The Committee then applies a Phased Grading process in making pass/fail determinations based on candidates’ total scores (Description and Grading of the California Bar Examination – General Bar Examination and Attorneys’ Examination). To pass the examination in the first phase of grading, a candidate must have a total scale score of 1440 or more out of 2000 possible points. Those with total scale scores after one reading below 1390 fail the examination. A candidate who receives a final score near the passing score after the First Read (Phase I) that is near the passing score will have all essay and Performance Test responses read a second time by a different grader (Phase II). Ratings assigned to responses on essay and

Performance Test questions are then averaged for Phase I and Phase II readings. If a candidate's averaged total scale score is 1440 or higher, the candidate passes the examination. If a candidate's averaged total grade is less than 1440 and there is no grading discrepancy greater than 10 raw points between the first and second readings of any essay or Performance Test question, the candidate fails the examination. Responses with grading discrepancies of more than 10 raw points between the first and second readings are read a third time by the Supervising EDG Team Member for that essay (Phase III). The Supervising EDG Team Member then resolves the discrepancy by assigning a third grade to the response, which is then used to calculate the final score and make a pass/fail decision. If the candidate's total scaled score after resolution grading is greater than 1440, the candidate passes the examination.

Finding: Grading of essay questions included on the California Bar Examination is based on the professional judgment of subject matter experts who undergo training and calibration in scoring candidate responses. Grading appears to be based on a general analytic scoring rubric that includes differential weights of essay components, as well as criteria for grading, both of which are developed during the first calibration meeting. The Committee provides a standard grading policy that instructs graders to consider the "overall quality of the answer, and whether the applicant has exhibited sufficient judgment, analytical ability, and knowledge of the subject matter involved in the question" (Committee of Bar Examiners' Policy Regarding Assignment of Grades to Written Assignments). These instructions further stipulate that grades are to be assigned on the content of the response and should not include other factors, such as quality of handwriting, accuracy of spelling or grammar, length of response, and inclusion of irrelevant matter. Scorers assign points in 5-point increments to essay components, up to the maximum weight provided on the scoring rubric.

However, the grading policy and scoring rubric do not appear to establish a clear link between the weighting of essay components or point assignments and competence for practice. The intended purpose of assessments used in licensure is to determine whether candidates possess the minimum level of competence required to practice safely and effectively upon entry into the profession. This purpose differs from assessments used in credentialing or other testing environments, which may assess competence at different levels or make relative comparisons among candidates (*Standards*, p. 169).

Standards states, "In the development and use of scoring rubrics, it is particularly important that credit be awarded for response characteristics central to the construct being measured and not for response characteristics that are irrelevant or tangential to the construct" (p. 56). As such, scoring procedures used in licensing examinations should be criterion-referenced. That is, grading specifications and procedures used for the California Bar Examination should be based on clearly defined criteria of what constitutes minimum competence to practice in the profession. A clear rationale for any differential weighting of essay components should be provided and should be based solely on these criteria. Further, a clear rationale should also be provided for incremental

point assignments made on essay components that are based on the objective criteria of minimum competence. Grading criteria, component weights, and incremental point assignments should not be made based on any factor or level of proficiency beyond minimum competence.

To increase standardization and comparability of equivalent forms, these criteria should be specified as part of test specifications used during development processes. To the extent possible, scoring rubrics based on these criteria should be developed and evaluated prior to administration of assessment measures. Further, these competence criteria should be objectively and consistently applied to similar examination content across administrations.

Finding: After the initial calibration meetings, where graders work together to score essay questions, graders work independently to assign grades to candidate responses.

Finding: Grader agreement and consistency is evaluated after the first calibration meeting and again mid-grading cycle. Evaluations of agreement conducted after the first calibration meeting appear to be based on the ratings of 25 select candidate responses (July 2019 California Bar Examination Tentative Grading Statistics). One method of assessing rater agreement is based on the rank order of a selection of candidate responses from highest to lowest. Another check of rater accuracy is performed mid-cycle, but the level of analyses performed is not specified.

Finding: Grading processes for the California Bar Examination include an established method for resolving scoring discrepancies. This method of resolving discrepancies appears consistent with technical guidelines, provided it takes into account previous judgments made by subject matter experts who serve as graders.

Finding: Scores on the California Bar Examination are converted to the same measurement scale as the MBE in deriving a final total score (Description and Grading of the California Bar Examination – General Bar Examination and Attorneys’ Examination). The purpose of scaled scores is to account for form difficulty, to ensure that scores across forms hold the same meaning, and to ensure fairness among candidates. Scaled scores accomplish this by making statistical adjustments where small variances in difficulty occur. However, forms can be considered equivalent only to the extent that they measure the same content in the same manner and have relatively similar statistical characteristics (*Standards*, p. 95). The California Bar Examination includes five essay questions and a Performance Test question that are drawn from 13 different content areas. The content areas selected appear to vary across administration. In addition, each of the essays appears to differentially weight components in a nonstandard manner. Additional evidence should be provided to establish the equivalency of forms used on different administrations of the California Bar Examination.

RECOMMENDATIONS

While evaluating essay development and scaling procedures is outside the scope of this project, the following five recommendations are offered for further consideration:

Recommendation: Clarify how the weights of different parts of essay questions are derived. These weights should be clearly linked to the criteria of minimum competence for entry-level practice. To the extent possible, weights should be standardized and assigned as part of development specifications. In addition, documentation should be provided that links point assignments with the criteria of minimum competence and not a higher level of proficiency.

Recommendation: Evaluate the sufficiency of scaling techniques in equating forms across administrations. Forms intended to measure similar constructs across different administrations should include the same content, be written at relatively similar levels of difficulty and complexity, and possess similar statistical characteristics. Since the California Bar Examination includes five essay questions and a Performance Test question that are drawn from 13 content areas, additional evidence of equivalency of forms should be provided.

Recommendation: Evaluate the potential impact of construct-irrelevant variance associated with essay and Performance Test format in measuring minimum competence for entry-level practice (e.g., administration time).

Recommendation: Additional documentation should be provided to describe how scoring procedures and rubrics relate to the intended purpose of assessing minimum competence for licensure. Scoring procedure or decision models used for licensing examinations should be based on clearly defined criteria of minimum competence rather than relative judgments of proficiency.

Recommendation: Remove Phase III of the scoring process and implement a procedure to resolve differences in two phases. Use of minimum competence criteria can assist resolution.

CONCLUSIONS

The Committee has taken several steps to score the California Bar Examination in a manner that provides an objective evaluation of candidate performance. The grader training and scoring processes attempt to meet professional guidelines and technical standards. However, there are additional steps that can be taken to strengthen the performance standards by establishing a more direct link between scoring and the intended purpose (i.e., assessment of minimum competence required for licensure).

CHAPTER 7 | TEST SECURITY

STANDARDS

The following standards are most relevant to test security for licensing examinations, as referenced in the *Standards*.

Standard 6.6

Reasonable efforts should be made to ensure the integrity of test scores by eliminating opportunities for test takers to attain scores by fraudulent or deceptive means (p. 116).

Standard 6.7

Test users have the responsibility of protecting the security of test materials at all times (p. 117).

Standard 8.9

Test takers should be made aware that having someone else take the test for them, disclosing confidential test material, or engaging in any other form of cheating is unacceptable and that such behavior may result in sanctions (p. 136).

Standard 9.21

Test users have the responsibility to protect the security of tests, including that of previous editions (p. 147).

“In licensure and certifications tests, such actions [disclosing confidential information] may compromise public health and safety” (*Standards*, p. 136).

“Whenever there are significant stakes associated with testing, there will be pressures to break the rules at both the test taker and test administrator levels” (Olson and Fremer, 2013, p. 15).

FINDINGS

Test Security – California Bar Examination

The State Bar website provides information regarding the following security-related topics:

- A candidate admittance ticket contains an applicant number, a file number, and an NCBE number. Candidates are required to show their ticket to the proctor when entering the test center.
- If items other than those allowed are brought to the test center, candidates are required to leave them at their own risk outside of the examination area. Candidates are not permitted access to those items while the examination is in session.

- Candidates are instructed to bring any permitted personal items (e.g., prescription medication) into the examination area in a small, clear plastic bag. Larger, permitted items (e.g., pillow) do not need to fit into the clear plastic bag.
- Signs will be posted showing seating assignments by applicant numbers located on the admittance ticket. Candidates must occupy the same seating space throughout the examination.
- Upon arrival at assigned seats, candidates are required to fasten and wear an identification badge during administration of the examination. The badge must be fastened so it can be easily seen by the proctors and may be required to be shown at any time.
- At some time during administration of the examination, a candidate may be required to provide fingerprints.
- During the examination, candidates are required to show a government-issued ID card that has their photograph (driver's license, California identification card, or passport), for the purpose of verification of identity. This ID card must be carried at all times during the examination.
- Candidates are not allowed to make notes from memory prior to the start of the examination session. If using a laptop computer, a candidate may not access documents on the computer after entering the secure examination area.
- Examination questions and used scratch paper may not be taken out of the examination area. After completing each session of the examination, all examination questions and used scratch paper must be placed in the envelope designated for that purpose.
- Timing devices (e.g., watches and clocks) brought to the examination test centers are to be used solely for the determination of the passage of time. They cannot have a digital display, must be absolutely silent, and cannot be larger than 4" x 4". Timing devices that are digital, programmable, or make noise are not permitted.
- Restrooms and water will be available at each test center. Candidates are instructed to use the facilities before taking seats to avoid missing the instructions prior to the start of each session. Extra time to use the restroom is not granted. Candidates are not permitted to use the restroom or leave seats during the last several minutes of the examination session. After time is called, candidates are not permitted to use the restroom or leave seats until all materials are collected and inventoried.
- If candidates leave the secured examination area during the session, they are not permitted to return for the purpose of completing the examination.
- At laptop computer test centers, laptop computers may not be disconnected from electrical outlets or removed from the examination room while the examination is in session, even if a candidate has completed the examination for that session. Candidates must wait until the examination session has concluded and candidates are dismissed before retrieving laptop computers.

If a candidate is suspected of cheating, proctors are instructed to notify staff but not to bring attention to the candidate or others sitting around the candidate. Incident reports are completed documenting the suspected cheating for investigation.

Failure to adhere to examination rules is considered a conduct violation and may result in a Chapter 6 Notice (Rules of the State Bar of California. Title 4. Admissions and Educational Standards. Division 1. Admission to Practice Law in California. Chapter 6. Conduct at Examinations).

Finding: Security procedures are rigorous. However, with the current administration format of the California Bar Examination, security vulnerabilities exist.

RECOMMENDATIONS

Recommendation: Consider revisiting the policy on candidate intervention in response to flagrant incidents of cheating. If a candidate is observed to be cheating and the incident is witnessed by a proctor, it is acceptable to excuse the candidate from the testing situation. The candidate should not be accused of cheating but rather informed that they will be contacted at a later date. An incident report should be filed and used to support a Chapter 6 Notice. In some instances, immediate intervention can be justified to prevent unnecessary exposure of test items or questions.

CONCLUSIONS

Given the findings, the test security policies, procedures, and protocols meet professional guidelines and technical standards but can always be strengthened.

CHAPTER 8 | COMMUNICATION TO CANDIDATES AND STAKEHOLDERS

STANDARDS

The following standard is most relevant to communication to candidates and stakeholders regarding licensing examinations, as referenced in the *Standards*.

Standard 9.15

Those who have a legitimate interest in an assessment should be informed about the purposes of testing, how tests will be administered, the factors considered in scoring examinee responses, how the scores will be used, how long the records will be retained, and to whom and under what conditions the records may be released (p. 146).

FINDINGS

As presented in the prior chapters, the State Bar provides detailed information about the California Bar Examination. The information covers registration to sitting for the examination through the process of releasing examination results.

In addition, the State Bar webcasts Committee of Bar Examiners' meetings and provides links to the following studies that support the California Bar Examination program:

- Report 1: Recent Performance Changes to the Bar Exam
- Report 2: Conducting a Standard Setting Study
- Report 3: Conducting a Content Validation Study
- Report 4: Performance Changes on the California Bar Examination

Finding: The State Bar and Committee is transparent in the development, administration, grading, and performance of the California Bar Examination. The Committee goes above and beyond in the level of detail shared with candidates and stakeholders while attempting to maintain the security of the examination process.

RECOMMENDATIONS

Recommendation: The State Bar and Committee should refrain from sharing the topics of the essay questions. An examination outline, resulting from the current practice analysis, should be made available for candidates preparing to take the licensing examination. It is important for the State Bar to remember that the goal for candidates is to prepare for safe and competent practice as an attorney, not to prepare question by question to pass the California Bar Examination.

CONCLUSIONS

Given the findings, communication to candidates and stakeholders meets professional guidelines and technical standards.

CHAPTER 9 | CONCLUSIONS

REVIEW OF THE ADMINISTRATION OF THE CALIFORNIA BAR EXAMINATION

Throughout this report, recommendations are offered to improve the administration of the California Bar Examination. In other words, strategies to strengthen the pass/fail decisions from the California Bar Examination (i.e., validity) are given. Validity is not an all-or-none concept. Rather, validity is about accumulating evidence to support pass/fail decisions. The goal should be to regularly evaluate examination development and administration processes to boost fairness, reliability, and validity.

The State Bar should be commended for their willingness to release information about processes and procedures for constructive evaluation. It is obvious that State Bar management and staff are committed to building and administering a fair, reliable, and valid examination program.

Recommendations discussed below are based on the cycle of examination validation (See Figure 1) and professional guidelines and technical standards.

1. A policy on examination validation should be formulated to guide future examination activities, including practice analysis, development, administration, scoring, etc. This policy should be based on professional testing guidelines and technical standards associated with licensure examinations. Distinction between academic testing and licensing testing should be clarified.
2. The (suggested) policy to be developed should state that a practice analysis should be conducted approximately every five years. Depending on changes in the profession, the analysis could be a modified technique, employing focus groups composed of subject matter experts instead of a full population survey approach.
3. Subject matter experts should be used throughout validation activities. Professional guidelines recommend that subject matter experts who participate in validation activities for licensing examinations be: (1) currently licensed and actively practicing in the field; (2) in good standing; (3) representative of the diversity of practice or specialty areas involved in the profession; and (4) representative of practice in California in terms of geographic location and years licensed, with an emphasis on entry level. Board members, deans, and educators are discouraged from participating in *examination development-related* activities because of conflict of interest concerns and undue influence. Instead, board members and deans may observe orientations and trainings, excusing themselves from the activities once underway. Educators may participate in job analysis studies. Those associated with professional associations or other vested interests should also be excluded from *examination development-related* activities (e.g., grading) to avoid unintentional bias or influence.

4. The resulting examination outline (i.e., critical competencies to be assessed on the examination) from the practice analysis should stimulate a discussion about the most fair, reliable, and valid tool(s) to assess entry-level competence to practice as an attorney in California. It is highly recommended that the State Bar consider reducing the essay portion of the California Bar Examination to standardize the process. It appears the California Bar Examination is functioning more like a college “final or comprehensive” examination, attempting to assess as much as possible in two days. However, the purpose of a licensure examination is to measure (within practicality) the most critical competencies in a fair, reliable, and valid manner.

This recommendation could also reduce possible security issues with extending test time across days in response to accommodations. And construct-irrelevant variance associated with test anxiety and logistics may be reduced.

It is also recommended that the State Bar consider removing the MBE from the two-day administration and utilizing CBT. This strategy would likely reduce the cognitive load on candidates and provide greater convenience for scheduling (e.g., local test centers, self-scheduling).

5. Although significant effort is made to construct, administer, and grade the essay and Performance Test questions in a reliable manner, the standardization across administrations is severely lacking. Topics, questions, and weighting should be consistent. Rubrics for scoring should be consistent and based on entry-level standards. Timing for answering the essay and Performance Test questions should be established based on studies rather than historical testimony. For example, how the weights of different parts of essay questions are derived should be clarified. These weights should be clearly linked to the criteria of minimum competence for entry-level practice. To the extent possible, weights should be standardized and assigned as part of development specifications. In addition, documentation should be provided that links point assignments with the criteria of minimum competence and not with a higher level of proficiency.
6. Competencies not measured in the essay or Performance Test questions could be evaluated in vignette-style questions administered via CBT.
7. Consider revisiting the policy on candidate intervention in response to flagrant incidents of cheating. If a candidate is observed to be cheating and the incident is witnessed by a proctor, it is acceptable to excuse the candidate from the testing situation. The candidate should not be accused of cheating but rather informed that they will be contacted at a later date. An incident report should be filed and used to support a Chapter 6 Notice. In many instances, immediate intervention can be justified to prevent unnecessary exposure of test items or questions.

Overall, the State Bar is meeting professional guidelines and technical standards in most areas. Without making any significant changes to the testing format, at a minimum, the grading or

scoring processes should be significantly revised. However, it is highly recommended that the State Bar revisit the essay, Performance Test, and MBE format. While it is tempting to place the full burden of preparedness to practice on the California Bar Examination, the process of licensure is a multi-hurdle process. Each hurdle should have its own reliability and validity, preparing the student for practice, with the final hurdle being the licensure examination. Schools have the responsibility to educate and prepare students for practice. Too often, the focus becomes passing the examination and not entry into the profession.

As the population of candidates increases, as well as its diversity, it is important to reassess testing formats/tools, especially given the advancements in technology and item structure. What was once thought to be the best approach may no longer be as suitable when meeting fairness, reliability, and validity.

Again, the State Bar should be commended for being transparent and opening its processes for analysis and constructive recommendations. DPPR appreciated the cooperation and collaboration from State Bar management and staff.

CHAPTER 10 | REFERENCES

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

California Department of Consumer Affairs. (2018). *Licensure Examination Validation Policy*. Sacramento, CA: Office of Professional Examination Services.

Downing, S. M. (2006). Twelve Steps for Effective Test Development. In S. M. Downing and T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 3-25). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice. (1978). *Federal Uniform Guidelines on Employee Selection Procedures*. Washington, DC: Author.

Olson, J.F. and Fremer, J. (2013). *TILSA Test Security Guidebook*. Washington, DC: Council of Chief State School Officers.

INTERVIEWS

1. Bolus, R. (2019, October 8). Telephone interview.
2. Campbell, T. (2019, November 1). Telephone interview.
3. Cummins, L. (2019, October 23). Telephone interview.
4. Doell, C. (2019, November 5). Telephone interview.
5. Herschkowitz, D. (2019, November 20). Telephone meeting.
6. Romero, B. (2019, October 15). Telephone interview.
7. Schwab-Wilhelmi, H. (2019, November 1). Telephone interview.
8. Soriano, M. (2019, October 15). Telephone interview.

EDG GRADING GUIDELINES BY QUESTION

1. Feb 2019 CBX Q1 Wills/Trusts/Community Property Grading Guidelines
2. Feb 2019 CBX Q1 CommProp chart with totals
3. Feb 2019 CBX Q1 CommProp GGL summary
4. Feb 2019 CBX Q1 CommProp Consensus Answer
5. Feb 2019 CBX Q2 Torts Final Grading Guidelines CONFIDENTIAL
6. Feb 2019 CBX Q2 Torts Final Group Answer
7. Feb 2019 Q3 Model Answer Real Property (Landlord/Tenant)
8. Feb 2019 Q3 Score Sheet V2
9. Feb 2019 CBX Q4 Evid CivPro Copy of Copy of Scoring Guide
10. Feb 2019 Q4 Evid CivPro Copy of Copy of Scoring Guide Final V2 3/23/19
11. Feb 2019 CBX Q4 Evid CivPro GGLs Final V9 3/23/19
12. Feb 2019 CBX Q5 PR Consensus Outline V3
13. Feb 2019 CBX Q5 PR Draft Consensus Outline Analysis V1

14. Feb 2019 CBX Q6 Performance Test – Draft Outline
15. Feb 2019 CBX Q6 Performance Test – Draft Consensus Answer
16. Mar 2019 Q5 PR Draft Consensus Outline & Analysis

CALIFORNIA BAR EXAMINATION GRADING PREPARATION

17. Scope of the California Bar Examination General Bar Examination and Attorneys' Examination
18. Description and Grading of the California Bar Examination – General Bar Examination and Attorneys' Examination
19. Bar Examination Grading overview for apprentice graders
20. July 2019 GBX Grader Correspondence address list – Q1

CALIFORNIA BAR EXAMINATION PHASED GRADING PROCESS

21. California Bar Examination: Phased Grading
22. Committee of Bar Examiners' Policy Regarding Assignment of Grades to Written Assignments
23. June 17, 2019: July 2019 California Bar Examination Grader Information and Materials
24. CONFIDENTIAL: July 2019 California Bar Examination Tentative Grading Statistics
25. Aug 2, 2019 Instructions for Pre-Grading Assignments
26. Aug 5, 2019 Memorandum July 2019 California Bar Examination – Selected Answers
27. Sept 9, 2019: Information and Instructions for Re-evaluating Answer Files in the Exam Grader Community

JULY 2019 GBX GRADER EVALUATION FORM

28. July 2019 Grader Interest Form
29. Grading the Calibration Bar Examination Grader Instruction I. Analysis and Outline for the Essay Question
30. Grading the Calibration Bar Examination Grader Instruction I. Analysis and Outline for the Performance Test
31. Grading the California Bar Examination Grader Instruction II. First Calibration Meeting
32. June 17, 2019: July 2019 California Bar Examination Phase III – Resolution Grading
33. July 2019 California Bar Examination Grader Instruction IV. Grading Procedures
34. Grading the California Bar Examination Grader Instruction V. Third Calibration Meeting

THE COMMITTEE OF BAR EXAMINERS OF THE STATE BAR OF CALIFORNIA TESTING ACCOMMODATIONS NOTICE

35. Important information and guidelines for applicants granted testing accommodations during administration of the July 2019 California Bar Examination
36. California Bar Examination, Workroom Manual Testing Accommodations
37. California Bar Examination, Workroom Manual Laptop/Writers

38. Religious Headwear Petition (screenshots of conversation/approval) received through AIMS
39. Laptop/Writers Bar Exam – Workroom Tub Forms
40. Laptop/Writers Bar Exam Staff – Representative Tub Forms

THE COMMITTEE OF BAR EXAMINERS OF THE STATE BAR OF CALIFORNIA TESTING ACCOMMODATIONS NOTICE JULY 2019

41. Testing accommodations test centers July 2019 California Bar Examination
 - a. Laptop instructions – Essays 1, 2, and 3
42. Testing accommodations test centers July 2019 California Bar Examination
 - a. Laptop instructions – Essays 4, 5, and PT
43. Testing Accommodations Bar Examination work room Tub Forms
44. Testing Accommodations Bar Examination staff-representative and workroom Tub Form
45. (2019 sample schedules)
46. More than double time (adjust total amount of time for each session according to time granted)
47. Time and one-half (over two days)
48. Time and one-half (three-day standard)
49. Time and one-half (over four days)
50. Time and one-third (over four-day extended)
51. Time and three quarters (all sessions)
52. Time and two-thirds (over four days)
53. Double time
54. Double time (over two days)
55. Double time (over four days)
56. Accommodations granted (over two days) – candidates must bring lunch
57. Accommodations granted (over four days) – candidates must bring lunch
58. Extra 1 hour and 15 minutes (two-day standard)
59. Time and one-third (two-day standard)
60. Time and one-third (three-day extended)
61. Accommodations made (over two days)
62. Accommodations made (over four days)
63. Extra one-half hour (over four days)
64. Memo re: February 2020 California Bar Examination

LEAD SECURITY, SECTION, AND GENERAL PROCTOR ORIENTATION

65. Proctor Guidelines Testing Accommodations
66. Orientation for Security Proctors – Monday
(Conducted by Lead Security Proctor)
67. Critical Task List Testing Accommodations Test Centers Security Proctor
68. Critical Task List Security Proctors Laptop/Writers (pamphlet)
69. Proctor Guidelines Laptop/Writers (Pamphlet)
70. Critical Task List Laptop/Writer Test Centers Section Proctor
71. Critical Task List Laptop/Writer Test Centers Floor Proctor
72. Orientation for Section Proctors
73. July 2019 Tips for Section Proctors

- 74. Critical Task List Testing Accommodations Section Proctor
- 75. General Proctor Orientation for Bar Examination Accommodations July 30, 2019
- 76. General Proctor Orientation for Bar Examination Accommodations July 29, 2019
- 77. General Proctor Orientation for Bar Examination Accommodations July 31, 2019
- 78. Guidelines for Lunches at TA Test Centers

ADMINISTRATION OF EXAMINATIONS EMERGENCY PROCEDURES (1994)

- 79. Staff Member in Charge and Staff Representative Manual July 2019 California Bar Examination
- 80. Staff Representative Daily Critical Task Reference Testing Accommodations
- 81. Staff Member in Charge (SMIC) Task List July 2019 Bar Examination Laptop/Writers July 29, 2019
- 82. Staff Representative Daily Critical Task Reference Writers/Laptops

THE COMMITTEE OF BAR EXAMINERS OF THE STATE BAR OF CALIFORNIA PREVIOUS TESTING ACCOMMODATIONS NOTICE FORMS FROM 2018

- 83. July 2018 Sample California Bar Examinations RA schedules
- 84. Time and one-third (2-day: Director approval required)
- 85. Time and one-half (All essays day 1) July 2018
- 86. Time and one-half (2-day proposed)
- 87. Note: 8:15 p.m. end time on Tuesday; 7:30 p.m. end time on Wednesday

(This page intentionally left blank.)

Attachment D

A REPORT ON

THE PHASED GRADING OF THE CALIFORNIA BAR EXAMINATION:

A PROFILE OF RECENT RESULTS AND

MODELING THE IMPACT OF ALTERNATIVE APPROACHES

Roger E. Bolus, Ph.D.

Research Solutions Group

February 27, 2020

A report prepared for the California Committee of Bar Examiners

EXECUTIVE SUMMARY

Since the early 1980's, grading of the California Bar Examination (CBX) written section has been conducted in a multi-stage (phased) process. While the decision rules defining each of the phases have changed over time, the fundamental structure has remained the same: provide additional gradings/reviews to the answers of examinees whose initial read scores are close but below the passing score. The benefit of the process is to ensure that those examinees do not fail as a result of errors in grading with an associated cost of prolonging release of results to all applicants and the public. The current study investigated how the phased grading process has fared since the introduction of the 2-Day CBX format and then conducted simulations to estimate the impact of various changes designed to hasten completion of grading.

During the first four administrations of the 2-Day CBX (July 2017 and 2018, and February 2018 and 2019), grading and scoring took place over three phases: Phase 1 in which all applicants answers are read and graded, Phase 2 in which those examinees scoring between 1,390 and 1,439.99 are graded a second time, and Phase 3 where the scores of Phase 2 failing applicants that are discrepant from the first grading are read for a 3rd time. The latter two phases required an additional 14 and 21 days of grading for the February and July administrations, respectively. Analyses showed that during those exams, close to 3,300 examinees out of 25,700 test-takers went into regrade; averaging about 970 examinees per July exam and 670 examinees per February exam. During the regrading, an additional 246 and 154 examinees passed, increasing the overall passing rate by an average of 1.45% in July and 1.65% in February. No particular subgroup of applicants (i.e., based on race/ethnicity, gender or type of law school attended) benefited from the regrade process more than another.

Analyses further demonstrated that those examinees whose scores were closest to the passing standard of 1,440 after their initial grading had a much higher likelihood of passing upon second grading. Across the four administrations, between 34% and 43% of examinees with scores of 1435 to 1439.99 after the 1st Phase, passed on regrade, while only 7% with scores in the 1,415 to 1,419.99 phase 1 score range passed. Across the four administrations, no applicant with a first read score of 1,400 or below (the bottom 20% of the regrade range) passed during the regrade. Only 8% (21 out of 261) making it into the 3rd phase (Reappraisal) passed. The overall CBX score reliability did not improve as a result of the incremental readings.

Based upon these results, alternative grading models were simulated. For each model, changes to passing rates, consistency with actual results, differential impacts on subgroups and changes in reliability were examined, along with the expected reduction in grading time.

- As a baseline, a simulation model (Model 1) that eliminated the regrade process completely (a method used by several state jurisdictions) resulted in a reduction of 21 and 14 days of grading, at a cost of dropping the average passing rates in July and February by 1.5% and 1.7%, respectively. Approximately 98% of test-takers would have been predicted to have the same outcome as they had in the actual grading, and no racial/ethnic nor gender subgroup was disparately impacted.
- A 2nd Model, eliminating the Reappraisal Phase of regrade only was estimated to result in a reduction of only 7 days of grading, but with a smaller impact on applicant outcomes (only a .1% drop in the overall pass/rate and 99.9% decision consistency rate).
- Finally, a 3rd model evaluated the estimated impact of *reducing the size* of the regrade range by 50% (i.e., 1415-1439.99). Based on current grading rates (i.e. answers read per grader per day), this model was expected to shorten grading by an average of 4

days on February exams and 8 days on July exams. The estimated impact on passing rates and decision consistency were similar to those of Model 2.

While undertaking the study, State Bar staff indicated that recent improvements in the automated data management systems could potentially provide improved management of graders and subsequently, the pace with which they read and score answers. Considering that graders might be able to read and score answers at a rate of 30 answers per day (Model 4), it was estimated that 19-20 days of grading (inclusive of the first phase) could be eliminated for July exams and 16-17 days for February. At this grading pace and the addition of two graders per question team (from 12 to 14; Model 5), an additional 3 to 4 days of grading time could be potentially eliminated from the current schedule; all gains realized without any adjustments to the parameters of the current phased grading process.

In conclusion, our calculations identified potential adjustments to the current phased grading system that could reduce grading time while having differing impacts on examinees' outcomes. However, we do need to point out that, as with other simulation type studies, we maintained an "et ceteris paribus" (i.e., all things held equal) set of assumptions during the analysis. That is to say, the modeling assumed that other factors would not change simultaneously as the changes were implemented. Thus, while the study identified several possibilities for modifying the phased grading process, it is not clear in practice whether, and to what degree any of these structural changes might interact with actual grading behavior. For example, it would not be unreasonable to think that there might be an inherent tendency for graders to become more lenient, if they believe that the size of the re-grade range is getting smaller, or eliminated completely. Also, it is not a sure bet that the high levels of reliability remain intact if the pace of grading were to be increased. Obtaining feedback from the grading teams and their leadership would be critical before any modifications are

implemented, as would be pre-planning for a systematic assessment of the impact after changes have been implemented.

I. INTRODUCTION & BACKGROUND

Since the early 1980's, grading of the California Bar Examination (CBX) written section has been conducted in a multi-stage (phased) process. While the decision rules defining each of the phases have changed over time, the fundamental structure has remained the same: provide additional gradings/reviews to the answers of examinees whose scores are close but below the passing score. The philosophy and intent behind the practice has been to minimize the chances of making a “false-negative” decision, i.e., failing an applicant who is truly worthy of passing. This approach to grading was based on early experimental research conducted by Klein (1980) and is a common grading practice used by some, but not all states administering a bar examination.

Currently, the *written* section of the CBX is graded in three phases. During the initial phase (Phase 1) each of the written answers (five essays and one performance task) for all examinees are graded by one reader from team of twelve, who has been trained on the specific question and has gone through three separate calibration sessions early in the grading process. This initial grading phase generally requires between 3 to 5 weeks to complete depending on the administration (February or July) and the number of applicants sitting for the examination. After completion of this first phase, a total score (on a 0 to 2000-point scale) is calculated for each examinee by combining the written section score and the Multi-State Bar Examination (MBE)¹. A second grading (Phase 2) is then conducted on the answers of examinees whose initial total scale score was below but close to the passing standard of 1,440 (1,390 to 1,439.99). This second phase is currently completed in one to two

¹ The multiple-choice section of the CBX, the Multistate Bar Examination (MBE), is scored by the National Conference of Bar Examiners (NCBE) and sent back to the State Bar of California for inclusion in calculations of an applicant's final total score. The written scores are placed on the same scale of measurement of the MBE. Both sections are equally weighted and added together to arrive at the total scale score.

weeks after the completion of Phase 1. A final review, known as Resolution grading (Phase 3), is subsequently conducted on any applicant's answers whose first and second read scores have differed by more than 10 points². The Resolution phase currently takes an additional week.

Even with the shortening of the CBX to a 2-Day format, the volume of answers that need to be graded and the amount of time required to complete the task has been substantial. Over the first four administrations of the 2-Day examination format (first used in July of 2017), over 153,000 applicant answers have been graded (averaging 49,000 on the July and 27,500 on the February exams)³. This volume and time required is by far and away, the largest and longest of any state jurisdiction in the U.S. By policy, the results of the CBX are not released until *all* grading is complete and final scores are assigned. As a result, California is the last state to release results to the candidates and the public during each administration. Thus, there are continuing pressures on the State Bar of California to hasten the process without sacrificing its quality and the reliability of the results.

Efforts are continually underway to respond to the pressure for earlier release of results. The move to a 2-Day examination format, for example, actually reduced the length of the written portion of the examination by one essay question and one performance task, and subsequently the number of answers that required grading. Procedural efforts have also been implemented. For example, as of July 2017, the size of each question grading team was expanded from eleven members to twelve, thereby decreasing the work load of each grader with end-of-examination analysis demonstrating little to no decrease in test reliability. Each grader is currently required to read the answers of approximately 450 to 800 applicants (depending on administration), Simultaneously, modifications

² The 3rd Phase of grading has changed multiple times over the years based on research findings. The process has included a 3rd reading and scoring of all answers by a single grader assigning a simple pass vs. fail decision to actually averaging each of the essay/performance task scores from each phase. The current method has been in place since 2007.

³ These counts exclude the number of answers that need to be read for out-of-state lawyers sitting for the 1-Day Attorney's examination.

to the weighting scheme of the written and MBE portions of the CBX have been introduced contributing to improvements in examination reliability. Technology advances in the management of examinee answers and their distribution to graders are also currently underway, with the intent of minimizing logistical problems and possibly hastening the grading.

As mentioned above, despite these structural changes that have previously and are currently taking place, California remains the last state in the U.S. to release its results, despite administering the examination at the same time as the rest of the country. It can be argued that some of the most critical decisions in an examinee's life rests upon their outcome on the CBX and the faster they can receive their results, the better. As a result, the State Bar of California and the Committee of Bar Examiners, in particular, are continually challenged to evaluate the impact of the current examination processes, both in terms of making accurate decisions regarding examinees outcome as well as hastening the release of results to examinees and the public. Continual re-evaluation of the grading process in general and more specifically, "Phased Grading" offers one potential avenue for improvement.

II. RESEARCH ISSUES

Research Questions. This investigation proceeded in two stages. In the first stage, we focused on gaining an understanding of the volume of activity in the phased grading process, who was impacted, and what were the outcomes. Specifically, the purpose was to initially provide answers to the following questions regarding phased grading:

1. How many applicants⁴ enter each of the three grading phases?

⁴ The terms "applicants", "examinees" and "test-takers" are used interchangeably throughout the report.

2. How many eventually pass and fail within each phase?
3. How consistent are the findings across administration? By February vs. July?
4. How many answer books are actually being read during the process and when?
5. Do the applicants entering the various phases systematically vary by any relevant demographic (e.g., gender) or classification (e.g., type of school attended)?
6. To what degree is the CBX reliability impacted by phased grading?
7. What impact does the grading team sizes and pace of grading have on the outcome?

And subsequently at a more granular level, the objective is to profile process and outcomes *within* the actual regrade phases:

8. How many applicants enter Phase 2 grading at different Phase 1 score ranges (1,390 to 1439.99)?
9. How many applicants within each of those ranges eventually pass?
10. How many answer books are being read within each of the score ranges?
11. Does the composition of applicants (e.g., percentage of minorities) remain consistent throughout the range?
12. Is the size of the Phase 2 regrade band still appropriate?

Based upon the findings to this first set of questions, the second stage explored the *expected* effect of modifying the protocols of the current regrade practices on examination outcomes. Understanding that any change would lead to some impact, the second stage examined various alternatives to the parameters of the current phased grading system and attempted to quantify the potential impact in terms of:

- Differences in passing rates

- Consistency in the individual pass/fail decisions that would be made
- Whether the alternative would have a disparate impact on any particular applicant subgroup
- The ultimate reliability of the examination
- The number of examinees who would be expected to go into a phase of regrading and the resulting number of answer books that would need to be read
- The expected net impact on the release of results to the examinees and the public.

The alternative grading models that were selected included some that were based on the initial analysis stage and others that the Admission staff of the Bar thought might be plausible alternatives based on current operational considerations and constraints.

Data and Methodology. To profile the phased grading process and subsequently model possible alternatives, the two most recent years of General Bar Examinations (GBX) were selected for analysis. This time period was selected because (a) The format of those exams were consistent (5 essays, a performance task, and the MBE), (b) the administration of the exams were all conducted over a 2-day period, (c) the number of grading phases and the number of graders that were involved remained constant and (d) there was an equal number of February and July administrations (2 each) which allowed for an assessment of the consistency of results both between and among administration periods.

For each examination, all examinee scores were extracted from the Bar's historical databases, including individual written essay and performance task scores for each grading phase. Written, MBE and Total Scale Scores were also collected for each grading phase along with final pass/fail

status. Only applicants with a full set of scores (i.e., 5 essays, a performance task, and an MBE) were included.⁵

Demographic characteristics of each test-taker (e.g., sex and racial/ethnic status) and the law school attended were also extracted from the historical data. For selected analyses, the latter variables were recoded into broader categories (e.g., ethnic minority vs. non-minority and ABA graduate vs. graduate from non-ABA approved school)⁶.

⁵ Modeling analyses conducted in the second stage, included Attorney examination data since test-takers required grading during each phase and would contribute to the length of time required to complete grading.

⁶ The law schools were aggregated into the respective category at the time that the applicant first took the examination.

III. RESULTS AND FINDINGS

A. Stage I: Profiling Phased Grading

1. Foundational Statistics. Table 1 presents the number of examinees entering each of the grading phases during each of the four examinations, along with related passing statistics. The associated passing statistics include the number passing during the given grading phase, the percentage of applicants entering that phase who pass the examination, and the relative percentage that passing applicants represented of the total number passing the GBX.

Table 1
Number of Applicants Entering Each Grading Phase
And Related Passing Statistics Associated
With Each Phase

	July 2017 (N=8,546)				July 2018 (N=7,943)				February 2018 (N=4,654)				February 2019 (N=4,574)			
<u>Phase</u>	<u>N</u>	<u>N</u> <u>Pass</u>	<u>%</u> <u>Pass</u>	<u>% of</u> <u>All</u> <u>Pass</u>	<u>N</u>	<u>N</u> <u>Pass</u>	<u>% Pass</u>	<u>Cum.</u> <u>% Pass</u>	<u>N</u>	<u>N</u> <u>Pass</u>	<u>%</u> <u>Pass</u>	<u>Cum.</u> <u>% Pass</u>	<u>N</u>	<u>N</u> <u>Pass</u>	<u>%</u> <u>Pass</u>	<u>Cum. %</u> <u>Pass</u>
<u>Phase 1</u> <u>Regrade</u>	8,546	4,105	48.0%	96.9%	7,943	3,178	40.0%	96.8%	4,654	1,218	26.2%	95.0%	4,574	1,369	29.9%	93.8%
<u>Phase 2</u>	1,008	124	12.3%	2.9%	937	98	10.5%	3.0%	644	61	9.5%	4.8%	695	87	12.5%	6.0%
<u>Phase 3</u>	68	7	10.3%	0.2%	89	8	9.0%	0.2%	65	3	4.6%	0.2%	39	3	7.7%	0.2%
<u>Total</u>	8,546	4,236	49.6%	100.0%	7,943	3,284	41.3%	100.0%	4,654	1,282	27.5%	100.0%	4,574	1,459	31.9%	100.0%

All applicants are graded in the initial phase. During this initial phase approximately 97% of the applicants who would eventually pass a July examination were identified. On the two February administrations, between 94% and 95% of passers were identified. The differences between July and February rates are directly attributable to the lower average scores obtained by examinees in February.

In terms of the regrade experience in July, fully 13% of the applicants taking each examination went into regrade phases 2 and 3, with a fairly comparable proportion of those applicants passing during each of the phases (12.2% and 10.3% in 2017; 10.5% and 9.0% in 2018). While the passing rates within each of those phases were identical; the incremental contribution to the overall passing rates were slightly different. On the July administrations, Phase 2 passes represented about 3% of the overall passing rate, while the final phase contributed only .2%. In the absence of a regrade process (all things held equal), the overall July passing rate would have dropped by an average of about 1.5%.

Roughly similar regrade results were observed on the February administrations, though a slightly higher proportion of the overall passing rate came out of the Phase 2 grading. This finding is a direct byproduct of the lower scores and overall passing rates on the February examinations. During the Resolution Phase of the regrade process, which currently requires an additional week in the grading sequence, fewer than 20 additional examinees passed the CBX (7 and 8 on the two July administrations and 3 on each of the February administrations) over the four administrations.

Table 2 provides an estimate of the number of applicants answers that were reviewed and graded by readers during the three phases. During the two years under study, over 173,400 examinee answers were graded, averaging about 55,380 on the July administrations and 31,840 on the February administrations. At an average cost of about \$3.25 per essay answer and \$3.75 per performance task answer, the total costs of grading averaged about \$185,000 for the July examinations and \$106,000 for the February exams (excluding the fixed costs of grader training and calibration). Of this amount,

the cost associated with regrading answer booklets was estimated to be approximately \$19,700 per July examination and \$13,900 per February examination, and an additional three to five weeks of grading.

Table 2
Number of Answer Booklets Read
During Each Grading Phase

Phase	July 2017 (N=8,546)		July 2018 (N=7,943)		February 2018 (N=4,654)		February 2019 (N=4,574)	
	<u>Test-Takers</u>	<u>Answer Books</u>	<u>N</u>	<u>Answer Books</u>	<u>N</u>	<u>Answer Books</u>	<u>N</u>	<u>Answer Books</u>
Phase 1	8,546	51,276	7,943	47,658	4,645	27,870	4,577	27,462
Phase 2	1,008	6,048	937	5,622	644	3,864	695	4,170
Phase 3	68	70	89	92	65	68	39	42
Total	8,546	57,394	7,943	53,372	4,645	31,802	4,577	31,674

2. Phased Grading and Applicant Subgroups. Some have argued that the phased grading process provides a greater opportunity for under-represented subgroups of examinees to pass the CBX. That is to say, given the current passing standards, there is a prevalent belief that a disproportionate number of applicants from these groups (e.g., ethnic minorities) have an opportunity to enter into regrade (and subsequently pass) since their scores tend to be lower than the “majority” groups. This hypothesis has not been previously studied.

To address this notion, we first calculated the proportion of examinees that each of several subgroups represented of the total population. We then compared those proportions to the proportions based on the sample going into the regrade process. We reasoned that the greater the relative

proportions differed from one another; the more credence could be given to the argument that the regrade had a differential effect by subgroups. We subjected the proportions to a χ^2 test to evaluate whether any of the observed differences were statistically significant, or alternatively could have occurred by chance alone.

Table 3
Subgroup Representation in
The Total Population vs Regrade Sample

	July							February						
	2017		2018		Average		Diff	2018		2019		Average		Diff
Group	All Applicants	Applicants In Regrd	All Applicants	Applicants In Regrd	All Applicants	Applicants In Regrd		All Applicants	Applicants In Regrd	All Applicants	Applicants In Regrd	All Applicants	Applicants In Regrd	
Race														
Non_white	53%	51%	50%	51%	52%	51%	-1%	50%	52%	47%	46%	48%	49%	1%
White	47%	49%	50%	49%	48%	49%	1%	50%	48%	53%	54%	52%	51%	-1%
Sex														
Female	52%	54%	53%	52%	52%	53%	1%	53%	54%	54%	54%	53%	54%	1%
Male	48%	46%	47%	48%	48%	47%	-1%	47%	46%	46%	46%	47%	46%	-1%
ABA?														
No	17%	17%	17%	16%	17%	16%	-1%	29%	23% *	27%	21% *	28%	22% *	-6% *
Yes	83%	83%	83%	84%	83%	84%	1%	71%	77%	73%	79%	72%	78%	6%
CA ABA Tier														
1st	32%	39% *	30%	34%	31%	36%	5% *	48%	49%	45%	43%	46%	46%	0%
2nd	29%	32%	32%	34%	30%	33%	3%	29%	28%	32%	35%	31%	32%	1%
3rd	39%	29%	38%	32%	39%	31%	-8%	23%	23%	23%	22%	23%	22%	-1%

* Percentages in population and sample are statistically different (p < .001)

Table 3 presents the group percentages for four separate examinee demographic and other classification variables: racial/ethnic minority vs. non-minority, gender, type of law school (ABA vs. non-ABA), and CA ABA law school tier. Within each group, the relative percentages of each group

are presented for both the overall sample (i.e., all examinees) and the subsample of examinees going into regrade. Results are presented by each of the four examinations separately, and combined July and February administrations. χ^2 tests comparisons resulting in statistically significant differences ($p < .001$) between the population and regrade percentages are marked with a “*”.

Inspection of Table 3 reveals that the demographic composition of examinees in regrade generally do not differ from all CBX takers. With respect to race/ethnicity for example, across the two July administrations, Whites represented approximately 48% of the test taking population, while non-whites represented the balance (52%). Equivalent proportions were observed in the regrade sample (49% vs. 51%). The relative difference of only 1% was non-significant ($p=.41$). Identical non-significant findings ($p=.56$) were observed on the February exam, though in the opposite direction⁷. With respect to gender, the relative percentages of females and males in the total population and regrade samples followed a strikingly similar pattern of non-significance ($p=.57$ and $.58$ for the two combined February and the two combined July examinations). Comparisons by type of law school were less clear. While graduates of ABA and non-ABA approved schools shared similar representation in the overall population and regrade samples on the July exams (a 1% difference, $p=.49$), there was a 6% difference on the February administrations ($p < .001$). That is to say, on the February examinations relatively speaking, 6% more ABA graduates made it into regrade. And, when law school tier was examined, a pattern of relatively more graduates from upper-tier schools made it into regrade on the July examinations, but not on the February administrations. We suspect that the statistically significant differences (though still relatively small in an absolute sense)

⁷ A separate set of analyses was conducted by individual racial/ethnic groups (rather than all non-white groups aggregated in a single minority group). These analyses were conducted for each administration. No subgroup was found to be over or under represented within regrade and none of the χ^2 comparison tests were statistically significant at $\alpha=.01$ level.

found in the various school groups may be a function of the composition of repeaters in the respective groups (e.g., less first time, out-of-state ABA school graduates in February relative to July).

A final set of analyses on subgroups (detailed data not presented) , found that once in regrade, no subgroup of examinees (i.e., racial/ethnic, males vs. female, ABA school graduates vs. others, or graduates from different law school tiers), passed at a statistically significant higher rate ($p < .01$) than any other subgroup. *Thus, in the main, the results from Table 3 suggest that the phased grading process, either in terms of entry into regrade or outcome of the regrade process itself, did not serve to enhance the probability of a favorable outcome for one group of applicants at the expense of another.*

3. Phase 2 Grading Range. As mentioned in the Introduction, the current regrade band extends from 1,439.99 down to 1,390.00. As best as can be gleaned from available State Bar documents, the 50-point regrade band was based on previous quantitative research, that suggested that no applicant outside of that range would achieve a passing score upon further review. That range also happens to be approximately one Standard Error of Measurement (SEM⁸). In theory, about 70% of the time, an examinee's true level of performance would be expected to lie somewhere between one SEM above or below the score that they earned.

The CBX Phase 2 regrade band has remained static for many years, despite recent changes in the configuration of the examination (i.e., number of written questions) and scoring algorithms (e.g., changing the weight assigned to the respective written and MBE sections). Therefore, a more detailed look at what has occurred specifically within the Phase 2 band, was warranted. Providing a more

⁸ The SEM is an index of the average amount of error in a score. It is based on the reliability of the examination and the score spread.

granular look within the regrade range may provide further insights into making adjustments and refinements to the process.

Table 4 presents statistics on the number of examinees entering Phase 2 grading based upon their Total score (MBE and initial written score). It also shows the total number of answers that have been read. The data in Table 4 differs from earlier tables in that the results are presented within ten 5-point score bands spanning 1390.00 to 1439.99. Results are presented separately for the two July and two February administrations.

Looking first at the July administrations, it can be seen that the number of applicants within each of the 5-point score bands do not vary much, ranging between 8% and 13% of the total number entering Phase 2, with the modal percentage around 10%. A χ^2 test indicated that any differences between the two July administrations in terms of the score distributions were due to chance alone ($P>.01$). The February administrations showed slightly more variation between score bands (6% to 14% in 2018 and 9% to 11% in 2019), though the χ^2 test yielded similar non-statistically significant results ($P>.01$). Taken together, the results suggest that the current Phase 2 regrade band occurs within the portion of the overall distribution of scores where there is a fairly uniform percentage of applicants throughout the range, i.e., the percentages are getting neither larger or smaller as the scores decrease.

While the relative proportion of applicants tend to be equivalent within each of the score ranges, the percentage within each band who earn enough score points upon regrade to pass, would be expected to vary. The reason for this being that those closest to the 1,440 pass-point would be expected (based on this initial score) to possess more knowledge and better skills than those at the

Table 4
Number and Percentage of Test-Takers Entering Phase 2 Grading

By Five-Point Phase 1 Score Range

	July 2017 (N=1,008 in Regrade)			July 2018 (N=937 in Regrade)			February 2018 (N=643 in Regrade)			February 2019 (N=695 in Regrade)		
	Test-Takers	Answer Books	% of Regrade	Test-Takers	Answer Books	% of Regrade	Test-Takers	Answer Books	% of Regrade	Test-Takers	Answer Books	% of Regrade
Regrade Band												
1435-1439	107	642	10.6%	91	546	9.7%	40	240	6.2%	68	408	9.8%
1430-1434	100	600	9.9%	99	594	10.6%	80	480	12.4%	72	432	10.4%
1425-1429	104	624	10.3%	96	576	10.2%	63	378	9.8%	73	438	10.5%
1420-1424	115	690	11.4%	101	606	10.8%	59	354	9.2%	78	468	11.2%
1415-1419	95	570	9.4%	86	516	9.2%	72	432	11.2%	70	420	10.1%
1410-1414	121	726	12.0%	71	426	7.6%	66	396	10.3%	65	390	9.4%
1405-1409	97	582	9.6%	119	714	12.7%	47	282	7.3%	75	450	10.8%
1400-1404	80	480	7.9%	100	600	10.7%	60	360	9.3%	61	366	8.8%
1395-1399	103	618	10.2%	79	474	8.4%	94	564	14.6%	66	396	9.5%
1390-1394	86	516	8.5%	95	570	10.1%	62	372	9.6%	67	402	9.6%
Regrade Total	1,008	6,048	100.0%	937	5,622	100.0%	643	3,858	100.0%	695	4,170	100.0%

bottom of the range.⁹ and would simply require fewer additional points to pass. By how much and how quickly knowledge/skills change as first read scores decrease is an empirical question. Statistics addressing this question are presented in Table 5 on the following page.

The data in Table 5 quantifies the differences by presenting the eventual passing rates (based upon the combined first and second grading) within each of the 5-point, first read score bands. To offer further perspective, Table 5 also presents the incremental and cumulative passage rates, *relative to the overall bar passage rates* for each of the four examinations. Inspection of the statistics in Table 5 indicate that, as expected, the percentage of examinees eventually passing the

Table 5

⁹ It should be noted that graders who are scoring the answers for a second time, do not see the initial score, and therefore have no knowledge, where within the 1390 to 1439.99 range the examinee's scores fell.

Passing Rates of Test-Takers Entering Phase 2 Grading

By Five-Point Phase 1 Score Range

1390 to 1439.99

Regrade Band	July 2017 (N=1,008 in Regrade)				July 2018 (N=937 in Regrade)				February 2018 (N=643 in Regrade)				February 2019 (N=695 in Regrade)			
	<u>N</u> Pass	<u>% of</u> Band	<u>% of</u> CBX Pass	<u>Cum.</u> <u>% all</u> CBX Pass	<u>N</u> Pass	<u>% of</u> Band	<u>% of</u> CBX Pass	<u>Cum.</u> <u>% all</u> CBX Pass	<u>N</u> Pass	<u>% of</u> Band	<u>% of</u> CBX Pass	<u>Cum.</u> <u>% all</u> CBX Pass	<u>N</u> Pass	<u>% of</u> Band	<u>% of</u> CBX Pass	<u>Cum.</u> <u>% all</u> CBX Pass
1435-1439	45	42.1%	1.1%	1.1%	31	34.1%	0.9%	0.9%	14	35.0%	1.1%	1.1%	29	42.6%	2.0%	2.0%
1430-1434	29	29.0%	0.7%	1.7%	19	19.2%	0.6%	1.5%	21	26.3%	1.6%	2.7%	20	27.8%	1.4%	3.4%
1425-1429	21	20.2%	0.5%	2.2%	25	26.0%	0.8%	2.3%	9	14.3%	0.7%	3.4%	13	17.8%	0.9%	4.2%
1420-1424	16	13.9%	0.4%	2.6%	9	8.9%	0.3%	2.6%	7	11.9%	0.5%	4.0%	15	19.2%	1.0%	5.3%
1415-1419	9	9.5%	0.2%	2.8%	6	7.0%	0.2%	2.7%	5	6.9%	0.4%	4.4%	4	5.7%	0.3%	5.6%
1410-1414	1	0.8%	0.0%	2.9%	3	4.2%	0.1%	2.8%	3	4.5%	0.2%	4.6%	4	6.2%	0.3%	5.8%
1405-1409	1	1.0%	0.0%	2.9%	3	2.5%	0.1%	2.9%	2	4.3%	0.2%	4.8%	1	1.3%	0.1%	5.9%
1400-1404	2	2.5%	0.0%	2.9%	2	2.0%	0.1%	3.0%	0	0.0%	0.0%	4.8%	1	1.6%	0.1%	6.0%
1395-1399	0	0.0%	0.0%	2.9%	0	0.0%	0.0%	3.0%	0	0.0%	0.0%	4.8%	0	0.0%	0.0%	6.0%
1390-1394	0	0.0%	0.0%	2.9%	0	0.0%	0.0%	3.0%	0	0.0%	0.0%	4.8%	0	0.0%	0.0%	6.0%
Regrade Total	124	12.3%	2.9%	2.9%	98	10.5%	3.0%	3.0%	61	9.5%	4.8%	4.8%	87	12.5%	6.0%	6.0%

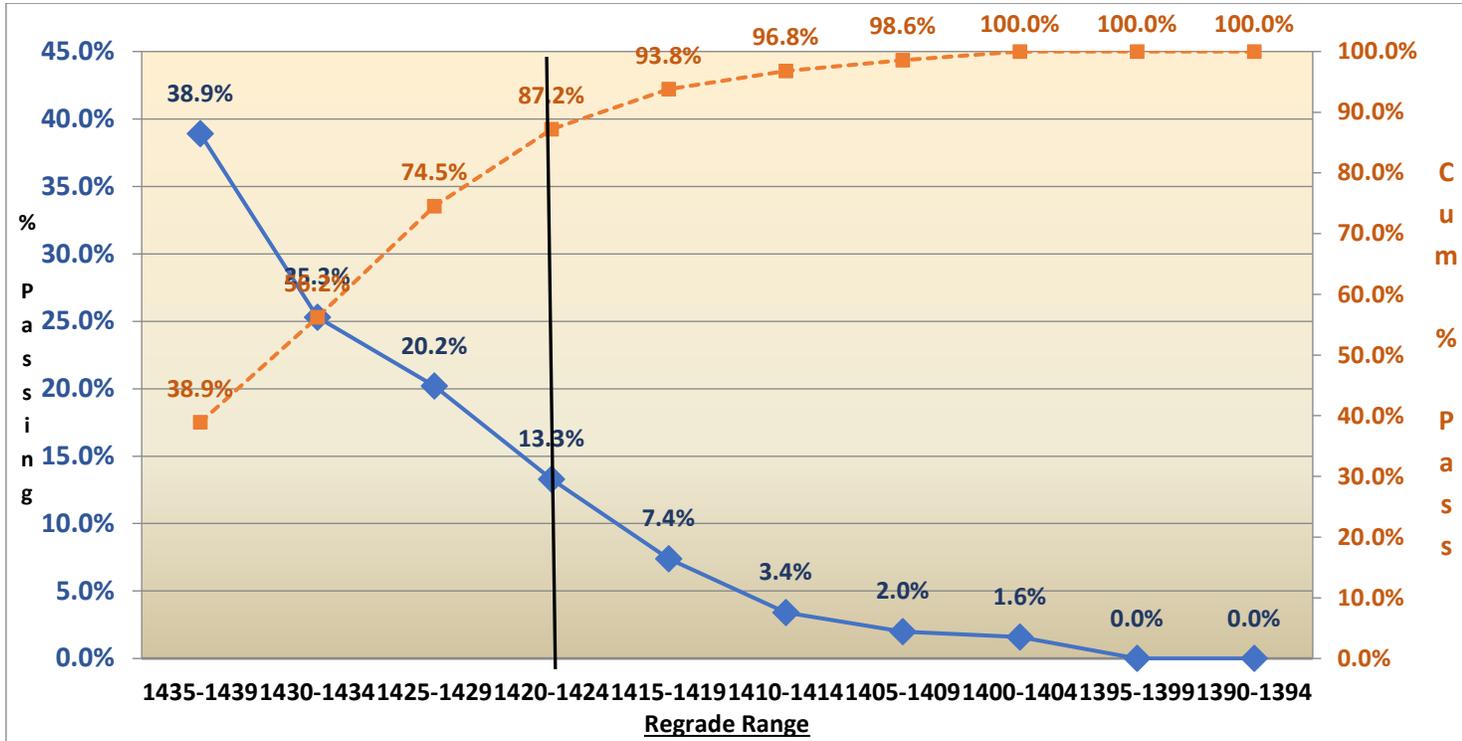
CBX after regrade decreases rapidly and systematically as the Phase 1 scores get smaller. For example, on each administration, between 34% and 43% of examinees who were within 5 points of passing after the initial scoring of the four examinations passed upon regrade. The passing percentages decreased to between 19% and 29% in the next two score (up to 10 Phase 1 scale score points lower). At the point in the Phase 1 score distribution when scores dipped to 1400 to 1404 (35 to 40 points below the actual passing standard), only 5 additional examinees eventually passed, representing less than .1% of the overall passage rate. *Across the four examinations under study, no applicant with a first read score of less than 1,400 passed during the regrade process.*

As demonstrated earlier, the impact of the regrade is relatively small, with about 3% of the passing rate in July and between 5% and 6% of the passing in February occurring during the process. The data in Overall, the data in Table 5 suggests that the majority of that impact of regrade tended to occur for examinees earning an initial score of 1,420 or higher on Phase 1 scores. Fully, 90% and 86% of the additional passers on the July examinations, and 83% and 89% on the February examinations came from above that score point. In the 30-point score range below that Phase 1 score, an additional 2,300+ examinees had over 11,600 of their answers regraded, only 47 of which would pass across the four administrations. Figure 1 on the following page further illustrates the impact on CBX passage rates throughout the Phase 2 grading range, highlighting the decreasing passage rates and the slowing cumulative rates at the 1,420-score point.

4. Improvement in Overall Test Reliability. One objective of the regrade process was to improve the reliability of scores, especially for those around the passing score of 1,440. Traditional reliability statistics reported on the examination have been based on composite written scores at the end of Phase 2 grading, which is the result of the double gradings for those in Phase 2 and the single grading for those who failed or passed at the end of the first grading phase. If overall test reliability has been truly enhanced as a result of the regrade process, then we might expect some increase in reliability as the additional gradings are added. To evaluate this assumption, we re-recomputed the overall Written test reliability using the scores available at the end of each phase¹⁰ and then compared the results. The resulting reliabilities are presented in Table 6 on page 23.

¹⁰ Cronbach Alpha was used to estimate the reliability at each phase.

Figure 1
Passing Rates of Test-Takers Entering Phase 2 Grading
By Five-Point Phase 1 Score Range
July 2017 through February 2019



Results in Table 6 conclusively show that the additional gradings had no impact on the overall reliability of the Written section of the CBX (and subsequently the overall CBX). For each administration, the reliability estimates are identical whether a single grading is used (Phase 1), a combination of single and double gradings (Phase 2) or a resolution grade is applied (Phase 3). The results are consistent for both February and July administrations. Thus, while the additional gradings may have somewhat improved the accuracy of scores for examinees entering regrade, it had no overall effect on the examination. We suspect that this is due to the consistency in graders' standards resulting from the extensive calibration process.

Table 6

**CBX Written Test Reliability
Based Upon Applicant’s Essay and Performance Test
Scores at the End of Each Grading Phase**

	<u>Phase 1</u>	<u>Phase 2</u>	<u>Phase3</u>
<u>July</u>			
2017	.73	.73	.73
2018	.76	.76	.76
Ave.	.75	.75	.75
<u>February</u>			
2018	.70	.71	.71
2019	.70	.71	.71
Ave.	.70	.71	.71

5. The Size of Grader Teams. Another factor contributing to the length of time required to release results is the number of individuals that are employed to perform grading of the Written portion of the CBX. As mentioned in the introduction, beginning with the first administration of the 2-Day format of the CBX, the size of the grading team assigned to read each essay and performance test during each of the first two grading phases was expanded from 11 to 12. One concern that was expressed before going to 12 graders per team was that the addition of more graders would potentially lower the reliability of the examination due to the anticipated difficulty of calibrating a larger team. According to State Bar staff, the addition of the single grader to each team was considered as a first step to possible further additions. Because the increase in grading team size corresponded to the shortening of the examination, until now it has not been possible to untangle the effects of the two simultaneous changes on test reliability. Isolating and studying the impact of team size on test reliability would lend support to the possibility of future grader additions.

To evaluate the potential effect, we extracted written scores from the four 3-Day examinations immediately preceding the July 2017 CBX (July 2015 and 2016, and February 2016 and 2017). We then, randomly selected five essay questions and one Performance Task score for each applicant to simulate a format consistent with the 2-Day format¹¹. We estimated the internal consistency reliabilities of the four examinations during the administrations where 11-member grader teams were being used, and then compared them to actual reliabilities of the four 2-Day examinations. All calculations were done on initial read scoring.¹² The results are illustrated in Table 7.

Table 7
Comparison of Written Test Reliability
11 vs. 12 Member Grading Teams
Based on Actual and Simulated 2-Day Exam Formats

Month	<u>11 Graders</u>			<u>12 Graders</u>		
	<u>1st Exam</u>	<u>2nd Exam</u>	<u>Ave</u>	<u>1st Exam</u>	<u>2nd Exam</u>	<u>Ave</u>
February	.691	.656	.674	.701	.696	.699
July	.766	.701	.734	.729	.757	.743

The results from Table 7 suggest that introduction of a 12th grader would not adversely impact reliability, and may even improve it. On the simulated 2-Day examinations, when 11-grader teams were used, the average written score reliabilities across the two February and July CBX were .674 and .734, respectively. On the actual 2-Day examinations, where 12 member teams were used, those averages were .699 and .743 respectively. While these findings are not based on direct comparisons

¹¹ We note that under the 3-Day format, applicants were given 3 hours (180 minutes) to answer the performance task.

¹² While it is true that questions were different on each of the exams, over the past 8 CBX exams, written score reliabilities have differed at most by .04 points.

(i.e., the questions were different on each examination, possibly effecting the reliability estimates), they are in the direction of suggesting that larger grading teams may be possible without sacrificing test reliability.

B. Stage II: Evaluating the Impacts of Alternative Phased Grading Models

The above findings point to opportunities for making further modifications to the current grading process that could potentially hasten the release of California's Bar Examination results. As referenced in the introduction, any modifications that would be made would naturally be expected to have impacts in a number of different areas. Therefore, clearly specifying alternatives and then quantifying their impact would be critical for making any policy changes. In this second section, we evaluate several alternative changes to the phased grading process and then statistically model their impact using results from the last two July and February examinations. The alternatives that are evaluated are based upon suggestions from State Bar staff as well as the results from the analyses conducted in the first part of the report. For each alternative model, we evaluate the estimated outcomes relative to the actual outcomes from the exam. Specifically, we present data on the following:

1. **Differences in passing rates.** First, we compare the actual passing rate to what the expected passing rate would be based on the specific changes in the model.
2. **Consistency in the individual pass/fail decisions.** Secondly, we assess the percentage of examinees that would have the same pass/fail status under a revised scoring as in the actual scoring. The higher the percentage, the less impact that the alternative model would have.
3. **Disparate impact.** We then determine whether the alternative model would have had a differential impact on the passing rates of different subgroups (e.g., gender or race). We do this by comparing the actual passing rates within each subgroup to the estimated passing rates and compare the differences using a χ^2 test of statistical significance. *A non-significant difference implies that there is no disparate impact.*
4. **Difference in Examination reliability.** The alternative models result in fewer gradings. Theoretically, fewer gradings would lead to lower reliability. Each of the models would **have a**

different number of gradings, and the objective of this analysis is to determine the degree to which each lowers the reliability and if so, by how much.

5. **Number of examinees and answers.** Each of the alternative models would either reduce the number of examinees who would have their answers subsequently graded, which in turn would reduce the number of answer books that are read. This analysis attempts to quantify these values and then compares them to the actual volumes that occurred in the original grading. The differences reflect the expected net savings.
6. **Grading Time Saved.** Since each of the proposed alternative models might be expected to reduce the number of examinees (and subsequently the number of answer books) entering regrade, they should result in a reduction in the length of the overall grading process. To estimate how much time the process might actually save, we calculated the actual average number of examinees graded per day by the team of 72 graders (i.e., 12 graders for each of 6 questions) during regrade. Across the four administrations, the grading teams averaged approximately 110 and 80 examinees/day in February and July, respectively¹³. This results in an average of about 9.2 and 6.7 books/day read by each grader during February and July examinations, respectively. We assume that the difference in the July and February numbers (30 examinees and 2.5 books/day) is due to the logistics required to manage the larger July test-taking population. We applied the 110 and 80 examinee/day rate as our factor in determining how many examinees would need to not be re-graded to save one day of grading time.¹⁴

We evaluated five different alternative models; three structural (i.e., those effecting changes in parameters of the grading process itself, and two grader-based (i.e., those effecting the activity and number of graders during each phase). For the first three models we estimate the

¹³ The estimates were inclusive of lawyers sitting for the written test only. They comprised about 13% of the total number of examinees in regrade.

¹⁴ We further adjusted calculations based on the knowledge that our analysis sample excluded examinees taking the lawyers examination.

effect on all outcomes, for the latter two, we focus only on the potential impact on hastening the release of results.

Model 1: Elimination of the Regrade Process. The concept of regrading is not universally accepted nor applied in practice. Several states have no regrade process at all (e.g., Hawaii), others have an extremely small range, routinely regrading less than .5% of examinees (e.g., Ohio), while others double grade all applicants (e.g. Delaware). From a psychometric standpoint, the National Conference of Bar Examiners initially advocated a no-regrade policy for states participating in the Uniform Bar Examination (UBE) process, but later revised their stance, allowing states to select their own process. Thus, the decision to have a regrade process at all is considered somewhat arbitrary. The first alternative model that we evaluated therefore was one in which phased grading is completely eliminated, and pass/fail decisions are based solely on the scores received on the initial grading¹⁵.

The results of these estimations are presented in Table 8. What we observe from the calculation in Table is that elimination of the regrade process would have resulting in fairly consistent decrease in the overall passing rates across the administrations of the CBX varying by between 1.3% to 2%. In terms of the overall consistency rates (i.e., the percentage agreement in outcome), the agreement rates, by definition, would be quite high, with the only changes being the 1.3% to 2.0% that originally passed in regrading, would have failed. Our calculations and subsequent evaluation of the statistical significance of observed differences also indicated that elimination of regrading would have had no disparate impact on either gender or racial/ethnic groups though a slightly disproportionate percentage of applicants from ABA-approved law schools would have been adversely affected. This latter finding was not consistent across all administrations.

¹⁵ We acknowledge that the simulation models that are presented do not capture any changes in grader behaviors that would possibly come along with an actual implementation of the policy. In the current example, if graders knew in advance that their scores were final, there may be a tendency to be more lenient in their grading standard.

Table 8**Model 1****Estimating the Impact of Modification of the Regrade Process:*****Elimination of Entire Regrade Process***

<u>Exam</u>	<u>Change in</u>	<u>Consistent</u>	<u>Disparate Impact on Subgroups</u>			<u>Change in</u>	<u>Examinees</u>	<u>Reduction</u>	<u>Overall</u>
	<u>Passing</u>	<u>Pass/Fail</u>	<u>Gender</u>	<u>Race</u>	<u>School*</u>	<u>Exam</u>	<u>No Longer</u>	<u>in Answer</u>	
	<u>Rate</u>	<u>Decisions</u>				<u>Reliability</u>	<u>in Regrade</u>	<u>Books</u>	<u>Days</u>
								<u>Read</u>	<u>Saved**</u>
Feb. 2018	-1.3%	98.6%	No	No	Yes	.01	709	3,932	14
Feb. 2019	-2.0%	98.0%	No	No	No	.01	734	4,212	14
Jul. 2017	-1.6%	98.5%	No	No	No	.00	1,076	6,118	21
Jul. 2018	-1.3%	98.7%	No	No	Yes	.00	1,026	5,714	21

* ABA vs Non-ABA law school graduates

** Inclusive of applicants sitting for the 1 Day (Lawyers) examination

Overall reliability of the examination was estimated to remain about the same (differing by no more than .01), and each of the examinees that went into either Phase 2 or 3 of the grading process (along with their 6 answers) would no longer require regrading. As a result, at the present rate that the readers are reading and scoring answers, between two (February) and three (July weeks) would be eliminated from the overall process if California opted to eliminate regrading altogether.

Model 2: Elimination of the 3rd Grading Phase-Resolution. Presently, an entire week is allocated to the 3rd Phase of grading for both February and July administrations. As seen in Table 1, the number of examinees sitting for the 2-Day examination that entered resolution grading during the last four examinations ranged from 39 (a February administration) to 89 (a July administration). Table 3 provides the estimates of the impact of eliminating the Resolution phase.

Table 9

Model 2

Estimating the Impact of Modification of the Regrade Process:

Elimination of Phase 3-Resolution

<u>Exam</u>	<u>Change in Passing Rate</u>	<u>Consistent Pass/Fail Decisions</u>	<u>Disparate Impact on Subgroups</u>			<u>Change in Exam Reliability</u>	<u>Examinees No Longer in Regrade</u>	<u>Reduction in Answer Books Read</u>	<u>Overall Days Saved**</u>
			<u>Gender</u>	<u>Race</u>	<u>School*</u>				
Feb. 2018	-0.1%	99.9%	No	No	No	.00	65	68	7
Feb. 2019	-0.1%	99.9%	No	No	No	.00	39	42	7
Jul. 2017	-0.1%	99.9%	No	No	No	.00	68	70	7
Jul. 2018	-0.1%	99.9%	No	No	No	.00	89	92	7

* ABA vs Non-ABA law school graduates

** Inclusive of applicants sitting for the 1 Day (Lawyers) examination

On each administration, the Resolution grading process serves to identify a small handful of applicants on each examination that should receive sufficient additional score points to achieve a 1,440 Total Scale Score, and subsequently pass the examination. Were the process eliminated, the annual CBX passage rate would be estimated to decrease by no more than .1%. The change would not impact the reliability of the overall examination, nor have any statistically significant effects on any of the subgroups under study. Based upon current grading protocols, it would be possible to reduce the length of time for grade release by one week if the Resolution phase was eliminated.

Model 3: Maintaining the Regrade Process-But Modifying Size of Regrade Band (1414.99-1439.99). The previous models evaluated elimination of the entire regrade process, or some portions of it. The analyses in Part 1 of this report demonstrated that as regraded examinees' 1st phase

scores decreased, the chances of their passing decreased as well, to the point where no one passed with a 1st phase score of less than 1400. The associated “cost” for extending the Phase 2 regrade band down to 1390 is to regrade examinees who have little to no probability of passing. Given the somewhat arbitrary nature of the 2nd Phase regrade band, we evaluated a model where the band would be cut in half, i.e., from 1415 to 1439.99. Under this model, examinees whose 1st Phase scores were between 1390 and 1414.99 would no longer enter regrade, but would be considered a “fail” at the end of the 1st Phase and the 3rd grading phase, Resolution grading, would be maintained. The results of the modeling are presented in Table 10.

Table 10

Model 3

Estimating the Impact of Modification of the Regrade Process:

Reducing Phase 2 Regrade Range – (1415-1439.99)

<u>Exam</u>	<u>Change in Passing Rate</u>	<u>Consistent Pass/Fail Decisions</u>	<u>Disparate Impact on Subgroups</u>			<u>Change in Exam Reliability</u>	<u>Examinees No Longer in Regrade</u> **	<u>Reduction in Answer Books Read</u> **	<u>Overall Days Saved</u> **
			<u>Gender</u>	<u>Race</u>	<u>School*</u>				
Feb. 2018	-0.1%	99.9%	No	No	No	.01	446	2,675	4.1
Feb. 2019	-0.1%	99.9%	No	No	No	.01	453	2,715	4.1
Jul. 2017	-0.1%	99.9%	No	No	No	.00	660	3,959	8.2
Jul. 2018	-0.1%	90.9%	No	No	No	.00	629	3,772	7.9

* ABA vs Non-ABA law school graduates

** Inclusive of applicants sitting for the 1 Day (Lawyers) examination

From the results in Table 10, we can see that reducing the regrade range for Phase 2 by 50%, results in virtually all examinees (99.9%) having the same estimated as actual pass/fail decisions. Additionally, there would be an expected and consistent (i.e., across examination) .1% decrease in

the overall passing rate (as compared to the actual rates). The modified rules would not be expected to have any differential impact on any of the subgroups, nor the overall examination reliability. The net effect would have been to shave about four days off of the regrade process in February and about eight days on the July administrations¹⁶. None of the examinees who would have failed under the modified process, but originally passed during regrade, entered the Reappraisal (Phase 3) process.

Model 4: Maintaining Regrade Process-Increase the “Grading Rate”. A 4th potential modification to the grading process would maintain all of the parameters of the grading phases (i.e., score bands and number of phases), but would alter with the activity of the graders themselves. Historically, due to the logistics of managing the distribution of examinee answers to graders and ensuring that graders did not begin subsequent grading phases until the preceding phase was complete, a fixed number of weeks has been allocated to completing each grading phase. With the advent of the 2-Day CBX administration format, the number of weeks in each phase was shortened somewhat and an additional grader was added to each team. For the subsequent two years, the time allocations were 21, 7 and 7 days respectively per phase for the February administrations, and 28, 14, and 7 days respectively per phase for the July administrations¹⁷. However, up to this point no formal consideration was given to the *rate per day* at which graders were expected to read and score answer books. After several discussions with State Bar staff with consideration of the new technologies currently in place, *a rate of approximately 30 answer books/day per grader was considered both desirable and potentially achievable*. Assuming this rate and the current team size of 12 graders per question, the total team of 72 graders could be expected to grade all answers for 360 examinees per

¹⁶ The 1-Day Attorney Exam accounted on average for an estimated additional 14% examinees in regrade in February and 11% in July.

¹⁷ The number of days allocated for the 1st grading phase include the time required for grader calibration sessions.

day, a rate that is significantly higher than what occurred over the last four administrations of the CBX.

To evaluate the potential impact on completion of the entire grading process (i.e., all grading phases), we assumed the completion rate of 360 examinees/day and applied that rate to the four administrations of the 2-day examination format¹⁸ to arrive at an Estimated (Est.) number of days that it would have taken to complete grading. We then calculated the difference between the Actual Time (Act.) to estimate the number of days that the overall grading process might be reduced (Diff.). The results are summarized in Table 11.

Table 11

Model 4

Estimating the Impact of Modification of the Regrade Process:

Improving the Rate with which Graders Read and Score Written Answers

Phase	July 2017			July 2018			February 2018			February 2019		
	<u>Act.</u> Days	<u>Est.</u> Days	<u>Diff.</u>									
Phase 1	30	26	-4	30	24	-6	23	15	-8	23	15	-8
Phase 2	14	4	-10	14	3	-11	7	3	-4	7	3	-4
Phase 3	7	3	-4	7	4	-3	7	3	-4	7	2	-5
Total	51	32	-19	51	31	-20	37	21	-16	37	20	-17

Table 11 presents the actual (Act.), estimated (Est.) days taken to complete grading for each Phase, along with the Difference (Diff.) for each of the four CBX administrations under the

¹⁸ Since only the team leader is responsible for grading during the 3rd Phase (i.e., Resolution), we applied a rate of 30 examinees per day.

assumptions that the graders could each read and score 30 answers per day¹⁹. The results in Table 11 suggest that having a more aggressive workload for the grading teams would not only have resulted in a significantly faster completion of not only the regrade period (Phases 2 and 3), but also reduction in the length of the initial phase. For the July CBX's, overall grading days would have been reduced by average of 19 days while February CBX grading would have been reduced by an average of 16 days. The size of the savings per grading phase, reflect the historical issues that State Bar staff needed to pre-establish fixed time periods per phase due to the logistics of physical book distribution.

While these calculations show that there are clear savings to be had by increasing the pace at which answers are read, it is unclear at this point what the impact on examination reliability might be if graders were held to a more aggressive “production” standard. For example, if such a standard were to be applied, it is possible that graders may feel excessively rushed to achieve daily or weekly quotas and subsequently make more errors in grading (e.g., by missing aspects in an examinees response that might contribute to a higher or lower score). If the State Bar was to consider this strategy, or some variation of it, some of the time savings and currently available technology, could be used to implement more “cross-grading” and real time, reliability checks.

Model 5: Maintaining Regrade Process-Increase the “Grading Rate” & Add Graders.

The last model that we considered is one that not only would increase the grading rate to 30 answer books per day per reader, but also expanding the size of the grading teams. According to State Bar staff, the most recent expansion of grader-team size from 11 to 12 members, from a logistical standpoint, was easily accommodated. The results from Table 7 above, suggested that the expansion

¹⁹ For the estimated figures, if a calculation resulted in a portion of a day (e.g., 2.1) the number was rounded up to the next whole number. Both Attorney and GBX exams are included.

did not have a negative impact on the overall examination reliability and might have even improved it. These findings leave open the possibility of expanding the teams further.

Another alternative for adjusting the Phased grading process is to not only increase the number of answer “booklets” that graders would read and score, but to also increase the size of the grading team. We modeled this approach by using the same calculations as Model 4, but adding another 2 graders to each team for a total of 14 graders per team for a grand total of 84 graders. This configuration would result in an estimated 420 applicants’ written tests being read and scored in a single day (as opposed to 360 applicants under the current team size. The results of this alternative (Model 5) is presented in Table 12.

Table 12

Model 5

Estimating the Impact of Modification of the Regrade Process:

Improving the Rate with which Graders Read and Score Written Answers &

Increasing the Number of Graders to 14 per Team

Phase	July 2017			July 2018			February 2018			February 2019		
	<u>Act.</u> Days	<u>Est.</u> Days	<u>Diff.</u>									
Phase 1	30	22	-8	30	21	-9	23	13	-10	23	13	-10
Phase 2	14	3	-11	14	3	-11	7	2	-5	7	2	-5
Phase 3	7	3	-4	7	4	-3	7	3	-4	7	2	-5
Total	51	28	-23	51	28	-23	37	18	-19	37	17	-20

Comparing the results of Table 11 and 12, we can see that the addition of two graders per team (12 in total), would have resulted in an estimated additional savings of about 3 more days per administration, with the largest impact coming during Phase 1 grading. No additional savings would be realized in Phase 3, since all grading is performed by the team leader.

Model Comparisons. If any of the various alternative models that we evaluated were implemented, it was estimated to result in changes to both CBX outcomes and the speed with results could have been released. However, each alternative resulted in different “costs” and “benefits”. To allow for a direct comparison of those costs and benefits, Tables 13 and 14 summarize the estimated impact of each alternative, averaging across the two July and February administrations separately. Table 13 focuses on the changes in the outcomes for the models that changed the “structural” parameters of the grading phases, while Table 14 focuses on the estimated savings (in days) of all models, including those related to grader management.

Table 13

Summary of Estimated Impact of Alternative Models on CBX Outcomes

<u>Modification</u>	<u>Difference in Pass Rate</u>		<u>Consistency in Pass/Fail</u>		<u>Differential Subgroup Impact*</u>		<u>Written Section Reliability Change</u>	
	<u>February</u>	<u>July</u>	<u>February</u>	<u>July</u>	<u>February</u>	<u>July</u>	<u>February</u>	<u>July</u>
	A. Structural							
1. Eliminate Regrade	-1.7%	-1.5%	98.3%	98.6%	No	No	0.01	0.00
2. Elimination Resolution	-0.1%	-0.1%	99.9%	99.9%	No	No	0.00	0.00
3. Shrink Regrade Range	-0.1%	-0.1%	99.9%	99.9%	No	No	0.01	0.00

* Gender and Racial/Ethnic groups only

Comparing the outcomes of the three models we see that their implementation would have had little to no estimated effect on key subgroups or the overall reliability of the examination.

Elimination of the Resolution (Phase 3) and reducing the size of the regrade range would be projected to have comparable effect on lowering the passing rate by only .1% (lowest of the four models) resulting in the highest consistency with actual examination results for the four administrations (99.9% of examinees with the same status). Complete elimination of the regrade process, all other things held equal, would have been estimated to lower the passage rates the most -by 1.7% and 1.5% on the February and July examinations, respectively.

These estimated impacts need to also be evaluated within the context of the savings in overall grading that each model would yield. Table 14 contains the comparisons for both the regrade phases only, as well for the entire grading process.

Table 14
Summary of Estimated Days of Grading “Saved” Under
Alternative Phased Grading Models

Modification	Regrade Only		All Grading Phases	
	February	July	February	July
A. Structural				
1. Eliminate Regrade	14.0	21.0	14.0	21.0
2. Elimination Resolution	7.0	7.0	7.0	7.0
3. Shrink Regrade Range	4.1	8.1	4.1	8.1
B. Graders				
4. Increase Grading Rate	8.5	14.0	16.5	19.5
5. Increase Grading Rate & N of Graders	9.5	14.5	19.5	23.0

Not unexpectedly, the larger the structural change, the greater the amount of expected savings in grading time (over current practice). The difference between February and July is a function of the differences in the current grading schedule. Maintaining the concept of a regrade process but

changing the regrade range, while having the smallest impact on outcomes (relative to current practice) would also be estimated to have the smallest savings. Modifying the management of graders (and doing nothing to change any of the parameters of phased grading) would be expected to yield savings that would fall in between the most and least invasive structural changes. (Note that these differences might have been even larger, but we assumed that the resolution process would still be carried out by a single grader per question.

Changes to the pace that graders would be expected to grade, and increasing the number of graders, would have the largest impact on Phase 1 grading. Implementation of either of the two alternatives involving change to the graders, would be expected to yield comparable savings to the models where the phased grading process was eliminated entirely.

IV. SUMMARY AND CONCLUSIONS

Over the past 20 years, multiple changes have been made to the CBX, among them the length of the test, the number of written essay and performance test questions that need to be answered, weighting of the respective test sections used in calculation of the final CBX scores, and most recently, the introduction of automation used in the distribution and recording of examinee scores. Many of the changes have been implemented in part to help hasten the delivery of results to examinees and the public. During that same period, the process of “Phased” grading of the written section has remained relatively unchanged.

In order to shed some light on what has been occurring in phased CBX grading a two-stage analysis was conducted. In Stage 1, the analysis profiled the phased grading process in terms of the number and characteristics of examinees experiencing each of the grading phases, along with the outcome of the process. The analysis also attempted to document the “costs” of phased grading in terms of the amount of time that incremental reviews (i.e., 2nd and 3rd gradings of the same examinee) added to the overall grading process. Using this information as a starting point, Stage II of the analysis then attempted to project what the expected outcomes might be if the process was modified in various ways. Those analyses also looked at the projected impact of altering the pace with which grader teams operate and the size of those teams.

The analyses in Stage I revealed that:

- Over the first four administrations of the 2-Day formatted GBX, close to 3,300 examinees out of 25,700 total test takers went into the regrade process, averaging about 970 and 670 for the two July and February administrations. The regrade process extended the release of results by 21 and 14 days in on the July and February administrations, respectively.

- Overall, during the regrade process an additional 246 and 154 applicants were identified as passing during the July and February administrations. These additional passing examinees resulted in increasing the overall GBX passing rates of the two July exams by 1.6% and 1.3%, respectively and the two February exams by 1.3% and 2.0%, respectively.
- No subgroup based on racial/ethnic status, gender, type or tier of law school attended were under or over-represented in the phased grading process, when compared to the overall population of examinees.
- A more granular inspection of the regrade range (set at 1390 to 1439.99 based on initial grading) indicated that roughly equivalent number of examinees entered the process throughout the range. However, the percentage passing dropped off rather quickly the further the initial score was from the 1440 pass-point. Between 34% and 43% with initial scores within 5 points of passing in the 1st Phase, passed on regrade across the four administrations, while an average of only 7% passed in 1415 to 1419 phase 1 score range. Across the four examinations, no applicant with a first read score of less than 1,400 (the bottom 20% of the regrade range) passed during the regrade process.
- Overall test reliability was unaffected by the additional grading as evidenced by the fact that the .75 and .71 written score reliability on the July and February examinations did not increase after all Phase 2 regrading was complete.
- An additional analysis indicated that counter to some early concerns, increasing the size of the grading teams from 11 to 12 members when the 2-Day format was initiated, had no material impact on the reliability of the GBX,

Based upon these findings in the 2nd analysis stage, a series of alternative phased grading scenarios were developed in order to evaluate their impact on exam and examinee outcomes and improvements in the time to release results. The models included various changes to “structural” parameters (e.g., changing the number of phases and/or size of the regrade range) and “grader” parameters (e.g., speeding up the pace of grading and/or adding graders). The calculations in the

models assumed that grading, scores and pass/fail decisions in the four examinations would have been made under the rules of the given model. Estimated results were then compared to actual results as a means of evaluating the model impact.

The models that were proposed in Stage II revealed that:

- Model 1. If regrade were eliminated completely, the overall passing rates on the examinations would be reduced by 1.3% to 2.0% depending on the administration. Approximately 98% to 99% of examinees would have received the same pass/fail determination, and there would be no disparate impact on any gender or racial/ethnic group. A very small, but statistically significant difference in passing rates was observed for graduates of ABA schools (compared to graduates of non-ABA schools) on two of the four administrations. Elimination of regrade resulted in no change in examination reliability and a reduction of 14 and 21 days in the grading processes of the February and July administrations.
- Model 2. Elimination of the Reappraisal process (Phase 3) alone was estimated to have the smallest estimated impact on passing rates (about .1%) and the highest consistency with the actual pass/fail decision (99.9%), accompanied by no corresponding impact on any subgroup of test-takers. These estimates were consistent for all four administrations and were expected to improve results reporting by seven days.
- Model 3. A 3rd model examined the impact of maintaining the regrade process, but cutting the Phase 2 regrade zone by 50% (1415-1439.99). The estimated impact of this model was to drop passing rates by .1%, achieve a 99.9% consistency rate in pass/fail decisions (both identical to the estimates of eliminating Reappraisal), while simultaneously showing no disparate impact on any subgroup and maintaining the same levels of reliability. This model resulted in estimated average savings of 4 days and 8 days of grading for the February and July administrations, respectively.

- Models 4 and 5. The final two models maintained all regrade rules currently in place but increased the pace with which graders evaluated and scored answer books (30/day; Model 4) and also added two members to each team (14 members/team; Model 56). Both simulations had the effect of decreasing the length of not only the regrade process, but also the initial grade phase. Model 4 was expected to reduce the over grading process by 19 to 20 days for July and 16 to 17 days in February. Adding two graders to the team, while reading and scoring at the 30/day pace, was estimated to result in an extra 3 to 4-day savings in July and about the same in February.

Overall, the findings from the study suggest that there are clearly opportunities for shortening the grading process of the CBX, with the potential for releasing final examination results faster than the current process. Savings can be realized by either modifying the structural parameters of the phased grading, changing the number and behavior of the graders themselves, or perhaps the combination of the two. Not unexpectedly, the study also demonstrated that the more the change deviates from current practice, the greater the expected impact on the examination and examinee results.

It is also apparent that recent technology enhancements in the State Bar's data management systems may facilitate enhanced grader management. Since paper versions of examinee answers no longer need to be distributed, electronic versions may be released randomly to graders in "real time" (i.e., as soon as the graders complete grading on one answer, the system can release another). We would anticipate that an alternative allotment strategy of answers to the graders could possibly result in improved monitoring and management of graders progress, This, in turn, could potentially hasten completion of the overall grading process. Also, since modifications to the phased grading system itself may be negatively perceived by the public, considerations of these system-based improvements might be given priority in the quest to speed up release of grades.

As with other simulation type studies (e.g., those that pre-modeled at 2-Day examination format), the “et ceteris paribus” (i.e., all things held equal) assumption was in play. As stated in the presentation of the results, it is unclear whether and how any structural changes to the phased grading process modeled above might interact with grading behavior. It would not be unreasonable to think that there might be an inherent tendency for graders to become more lenient, if they believe that the size of the re-grade range is getting smaller, or eliminated completely. Also, it is not a sure bet that the high levels of reliability remain intact when the pace of grading is increased. Obtaining feedback from the grading teams and their leadership will be critical before any changes are implemented, as well as pre-planning for an assessment of the impact after implementation.

References

Klein, S. A. Comparison of a Single vs. Multi-Phased Grading System. A report prepared for the State Bar of California (1980).