



# The State Bar of California

---

## **OPEN SESSION AGENDA ITEM APRIL 2020 COMMITTEE OF BAR EXAMINERS ITEM O-200**

**DATE:** April 24, 2020

**TO:** Members, Committee of Bar Examiners

**FROM:** Lisa J. Cummins, Program Manager III, Examinations  
Christina Doell, Program Manager I, Examination Grading

**SUBJECT:** Approval of and Action on Report Prepared by the Committee's  
Psychometrician, as Part of the Implementation of the Appendix I  
Recommendation to Evaluate the Grading Process for the California Bar  
Examination

---

### **BACKGROUND**

The work of the State Bar's 2017 Governance in the Public Interest Task Force included a review of the State Bar's various volunteer sub-entities, one of which is the Committee of Bar Examiners (Committee), "to assess whether the structure of the sub-entities aligns with assigned tasks and appropriate oversight mechanism are in place." The Task Force's report provided background on the Committee and recommendations for further study. Consultants engaged by the State Bar recommended that the Committee conduct a "review and evaluation of its grading process."

In response, the Board directed staff to conduct an evaluation of the grading process for the California Bar Exam. Staff consulted with the Committee's psychometrician, Roger Bolus, Ph.D., and requested him to conduct an evaluation of the processes for grading of the California Bar Examination (CBX), with an emphasis on identifying efficiencies and best practices, and to prepare a report on his findings.

Attached is Dr. Bolus's report, "A Report on the Phased Grading of the California Bar Examination: A Profile of Recent Results and Modeling the Impact of Alternative Approaches," dated April 9, 2020. Dr. Bolus will also be presenting his report at the Committee's April 2020 meeting.

A. California Bar Examination

The California Bar Examination (CBX) is made up of the 2-day General Bar Examination (GBX) and the 1-day Attorneys' Examination (AX). The GBX consists of two sections: the written section including five essay questions and one Performance Test (PT), and the 200-item Multistate Bar Examination (MBE). The six written questions are developed in-house by the State Bar of California and the MBE is developed and owned by the National Conference of Bar Examiners (NCBE). The AX consists of the five essay questions and one PT from the General Bar Examination only. Attorneys who are eligible to take the AX are not required to take the MBE.

To pass the examination an applicant must have a total scaled score of at least 1440 points out of a possible 2000 points.

B. Graders and EDG Team

The State Bar maintains a pool of approximately 150 graders who are all active California attorneys. Each grading cycle, 12 graders are hired for each of the 6 grading teams (one team per written question on the bar exam), for a total of 72 graders. In addition, 3-4 backup and apprentice graders are hired. Backup graders are generally ones who have participated in at least one prior grading cycle either as a grader or as an apprentice. Apprentice graders have never graded before. While the backup and apprentice graders are required to do the pre-grading assignments and to attend all three grading calibration meetings, they are not called upon to actually grade unless one or more of the primary graders drops out.

After the administration of the exam, but prior to beginning the grading process, the graders go through a 3-week training and calibration process to ensure that the grading guidelines are being consistently and appropriately applied. Graders are assigned a variety of tasks including writing a detailed analysis to the question they are assigned, participating in multiple calibration meetings to ensure all answers are graded according to the same standards, and participating in "tentative" grading to determine the level of each group's calibration.

The six members of the Examination Development and Grading (EDG) Team have many years of prior experience as graders and also have expertise in examination question development. They assist State Bar staff in developing the written questions for the CBX and also supervise the grading teams for each exam.

C. Scaling

The Committee utilizes a grading procedure designed to ensure the level of difficulty of the examination remains unchanged from one administration of the examination to another. The statistical technique, called scaling, converts scores on the written section (essay questions and PT) to the same scale of measurement as the MBE. MBE raw scores are converted to scale scores to adjust the results for possible differences in average question difficulty across

different administrations. As a result of this step, a given MBE scale score indicates the same level of proficiency regardless of the administration of the examination on which it was earned. Converting the total written raw scores to the same scale of measurement as the MBE adjusts for possible differences in average question difficulty and grading across different administrations of the examination.

On the written section of the examination, applicants are initially graded on a basis of 700 maximum possible points. An applicant can earn up to 100 raw points on each of the five essay questions and up to 100 raw points on the PT. The PT points are doubled in the calculation of the total raw written points. The scores obtained on the written section of the examination are then translated to the 2000-point MBE scale. An applicant'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.

Scaling ensures the two portions of the examination carry the relative weights assigned to them: Written (50%) and MBE (50%).

Attorney applicants who took the Attorneys' Examination also have their scores on the written section placed on the same scale of measurement as general applicants, but, as they are exempt from the MBE, their pass/fail status is based solely on the written section.

#### D. Phased Grading

All written answers submitted by applicants who completed the examination in its entirety are read and graded at least once before pass/fail decisions are made. For those applicants whose scores after the first read (Phase I) are below but near the required passing score, all answer books are read a second time, and the scores of the first and second readings are averaged. The total averaged score after two readings is then used to make a second set of pass/fail decisions, providing there are no discrepancies of more than 10 raw points between the first and second read assigned grades on any question (Phase II). Any answers with discrepancies of more than 10 raw points between the first and second read assigned grades are read a third time before a third set of pass/fail decisions (Phase III) is made.

To pass the examination in the first phase of grading, an applicant must have a total scaled score (after one reading) of at least 1440 points out of 2000 possible points. Those with total scaled scores after one reading below 1390 fail the examination. If the applicant's total scaled score is at least 1390 but less than 1440 after one reading, all of the applicant's answers are read a second time by a different set of graders. If the applicant's averaged total scaled score after two readings is 1440 or higher, the applicant passes the examination. Applicants with no discrepancies of more than 10 raw points between the first and second read assigned grades on any question with averaged total scaled scores of less than 1440, fail the examination. Applicants with grading discrepancies of more than 10 raw points between the first and second read assigned grades on any answer, whose averaged total scaled score is less than 1440, will have those answers referred to the Supervising EDG Team Member of the grading team for that particular question for resolution of the discrepancy. The EDG Team Supervisor will assign a

resolution grade to the answer and that grade will replace the average of the first and second read assigned grades for that question. Scores are calculated again and if the applicant's total scaled score after resolution grading is 1440 points or higher, that applicant passes the examination. If the applicant's total scaled score after resolution grading is less than 1440 points, the applicant fails the examination.

For ease of reference, the Committee's phased grading policy can be summarized as follows:

**PHASE I (First Read):**

Total Scaled Score (50% written + 50% MBE) = 1440 and above [PASS] – out of possible 2000

Total Scaled Score (50% written + 50% MBE) = below 1390 [FAIL]

Total Scaled Score (50% written + 50% MBE) = 1390 to below 1440 **[go to PHASE II]**

**PHASE II (Second Read):**

First Read and Second Read raw essay grades are averaged; average grade becomes operant grade for recalculation of scaled score.

Total Scaled Score (50% written + 50% MBE) = 1440 or above [PASS]

Total Scaled Score (50% written + 50% MBE) = below 1440 [FAIL] \*\*

**\*\* But, if below 1440 AND grading discrepancy of more than 10 raw points between first and second read grades on any question, then**

**[go to PHASE III]**

**PHASE III / RESOLUTION (Third Read):** [to Supervising EDG Team Member]

Resolution Grade replaces the averaged grade for that question (resolution grade can be no lower than lowest of the first and second read grades, and can be as high as 100); resolution grade becomes operant grade for recalculation of scaled score.

Total Scaled Score (50% written + 50% MBE) = 1440 or above [PASS]

Total Scaled Score (50% written + 50% MBE) = below 1440 [FAIL]

END OF GRADING

## E. Grading Timeline

Historically, the time it has taken from the examination administration until the release of results has been 11 and 16 weeks for the February and July examinations, respectively. The three phases of grading, along with the large number of applicants that test in California (approximately 5,500 in February and 9,500 in July), are significant factors in the length of time it takes for release of the results.

These timelines are designed so that graders can expect to spend approximately three (3) hours reading 20-25 answers each day, or between 140 and 175 answers each week. These numbers are provided as guidelines only, as they vary based upon the number of answer files assigned, the length and legibility of each answer and the amount of time each individual grader needs to read and evaluate the material.

## **DISCUSSION**

Dr. Bolus's study regarding the effects on the California Bar Examination Phased Grading Process has produced several alternative grading models to explore options for reducing the length of time between administering the CBX and releasing results. Based on Dr. Bolus's report, the Committee may decide that it is appropriate at this time to implement changes to some of the Committee's grading policies, processes or procedures for the CBX. This would be in line with the following Examinations Subcommittee Goal:

Review and consider proposals on ways to reduce the time it takes to release results for the California Bar Examination. Review and consider report prepared by the Committee's psychometrician, as part of the implementation of the Appendix I recommendation to evaluate the grading process for the California Bar Examination. If appropriate, consider and decide whether to implement any changes to the Committee's grading policies, processes, or procedures.

Dr. Bolus's report is being provided to the Committee in its entirety and Dr. Bolus will also be presenting his report. If the Committee decides to make changes to its CBX grading policies, processes or procedures, staff has recommendations based on Dr. Bolus's report that they believe will most effectively accomplish the goals of reducing grading time while having minimal impact on examinee outcomes. Those recommendations are provided below.

Dr. Bolus used the examination data over the last four administrations of the CBX (July 2017 and 2018, and February 2018 and 2019) to analyze and project what the expected outcomes might be if the regrade process (Phase 2 and Phase 3) were modified in various ways. In the report, a series of five alternative phased grading scenarios were developed in order to evaluate their impact on the exam and examinee outcomes, while condensing the timeline to release results. The five models are as follows:

Model 1: Eliminate regrading (Phase 2 and 3) completely

Model 2: Eliminate resolution grading (Phase 3)

Model 3: Compress the regrade band (currently a 50-point spread between 1390-1439.999)

Model 4: Increase the pace of grading

Model 5: Increase the number of graders assigned to each question (currently 12 graders)

Dr. Bolus summarized the findings in his report as follows:

- 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 1<sup>st</sup> 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, Dr. Bolus developed a series of alternative phased grading scenarios 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 models that were proposed 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 Resolution process (Phase 3) alone was estimated to have the smallest estimated impact on passing rates (about 0.1%) 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 speed up results reporting by seven days.
- Model 3. A 3<sup>rd</sup> 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 0.1%, achieve a 99.9% consistency rate in pass/fail decisions (both identical to the estimates of eliminating Resolution grading), 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 5). 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.

Taking into consideration all of the above from Dr. Bolus's report, together with their real world experience with CBX grading in actual practice, State Bar staff propose and recommend that the Committee make all (or in the alternative, any combination) of the following changes to its current grading policies, processes and procedures:

1. Eliminate Phase III of grading (Resolution) – Bolus Model 2
  - This change would result in a reduction in grading time of 7 days for both the February CBX and the July CBX.

- Phase III of grading serves to identify only a small handful of applicants on each examination that could possibly receive sufficient additional score points to pass the examination.
  - This change is estimated to reduce the overall CBX passage rate by no more than 0.1%.
2. Compress the Phase II regrade band (currently 1390 to 1439.99) to 1400 to 1439.99 – Bolus Model 3, modified
- Dr. Bolus’s model looks at compressing the regrade band by 50%, i.e., instead of a 50-point band, it would be a 25-point band (1415 to 1439.99). Dr. Bolus’s model estimated an average reduction in grading time of 4 days for the February CBX and 8 days for the July CBX, with a resulting in an estimated reduction in the overall CBX passage rate by 0.1%.
  - Staff proposes a modification of Dr. Bolus’s Model 3 regrade band, so as to create a 40-point regrade band (1400 to 1439.99) instead of a 25-point band. Dr. Bolus’s report demonstrated that, as regraded applicants’ first read (Phase I) scores decreased, the chances of their passing decreased as well, to the point where no one passed with a first read score of less than 1400.
  - Since Dr. Bolus’s report did not model a 40-point regrade band, it is unclear what the resultant impact of this modification would be on both grading time and pass rates. However, since the staff-proposed regrade band is more generous to applicants than the one modeled by Dr. Bolus, it would be expected to result in a reduction of the number of grading days from the current 50-point regrade band, but less than the number of estimated days saved in Dr. Bolus’s 25-point regrade band model. Similarly, because of the broader staff-proposed regrade band, the estimated reduction in the overall CBX passage rate would be something less than 0.1%.
3. Add two graders to each CBX question grading team – Bolus Model 5
- This change was estimated by Dr. Bolus to result in an extra 3 to 4-day savings in grading time for the July CBX and about the same for the February CBX. However, in reaching his conclusion, Dr. Bolus assumed that graders would be reading and grading answers at a pace of 30 per day (Model 4), which State Bar staff is not recommending at this time. Therefore, the addition of two graders would be expected to result in grading time saved, but perhaps not to the extent modeled in Dr. Bolus’s report.



The concern that staff have with Dr. Bolus's Model 4 to increase the speed at which the graders do their work is that, by compacting this timeline, we may risk graders feeling rushed and/or changing their normal grading behavior, whether they are aware of it or not. Dr. Bolus acknowledged this issue when he wrote in his report:

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 examinee's response that might contribute to a higher or lower score).

All CBX graders are practicing attorneys who are expected to have primary work duties outside of grading the bar examination. The current grading timeline is designed so that it provides graders the time necessary to give each applicant answer due consideration, while balancing the risk of grader fatigue setting in. The Committee uses a holistic, rather than a strictly point-based, grading process. Grading speed tends to be slower at the beginning and gets faster as more applicant answers are read.

Staff proposes that, instead of immediately requiring that graders read answers at an average speed of 30 answers per day, staff will solicit feedback from graders and then gradually increase grading speed, all the while closely monitoring the effects of any change on grader behavior, grader compliance, exam reliability, etc.

## **RECOMMENDATIONS**

It is recommended that the Committee receive and file Dr. Bolus's April 9, 2020 report and authorize its publication.

It is further recommended that the Committee adopt the following staff recommendations and direct staff to implement the following changes to its grading processes for the California Bar Examination as soon as practicable:

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

## **PROPOSED MOTION**

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

Move that the Report on the Phased Grading of the California Bar Examination: A Profile of Recent Results and Modeling the Impact of Alternative Approaches, prepared by Roger Bolus, Ph.D. and dated April 20, 2020, be received and filed and that the report be authorized for publication.

Further move that staff is instructed to implement the following changes to the Committee of Bar Examiners' grading processes for the California Bar Examination as soon as practicable:

- (1) Eliminate the Phase III (Resolution) grading;
- (2) Compress the regrade band for Phase II grading from 1390 – 1439.99 to 1400 – 1439.99; and
- (3) Add two graders to each California Bar Examination question grading team.

**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

April 9, 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 (currently 1440). 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 3<sup>rd</sup> 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 1<sup>st</sup> 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 3<sup>rd</sup> phase (Resolution) 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 2<sup>nd</sup> Model, eliminating the Resolution 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 0.1% drop in the overall pass/rate and 99.9% decision consistency rate).
- Finally, a 3<sup>rd</sup> 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 the study was being undertaken, 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 Multistate Bar Examination (MBE)<sup>1</sup> score. 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 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

---

<sup>1</sup> 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 combined to arrive at the total scale score.



second read scores have differed by more than 10 points<sup>2</sup>. 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)<sup>3</sup>. 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 to the weighting scheme of the written and MBE portions of the

---

<sup>2</sup> The 3<sup>rd</sup> Phase of grading has changed multiple times over the years based on research findings. The process has included a 3<sup>rd</sup> 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.

<sup>3</sup> These counts exclude the number of answers that need to be read for out-of-state lawyers sitting for the 1-Day Attorneys' Examination.

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 applicants<sup>4</sup> enter each of the three grading phases?
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?

---

<sup>4</sup> The terms “applicants”, “examinees” and “test-takers” are used interchangeably throughout the report.

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 State 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.<sup>5</sup>

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)<sup>6</sup>.

---

<sup>5</sup> Modeling analyses conducted in the second stage, included Attorneys' Examination data since test-takers required grading during each phase and would contribute to the length of time required to complete grading.

<sup>6</sup> 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>
<b>Phase 1</b>	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%
<b>Regrade</b>																
<b>Phase 2</b>	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%
<b>Phase 3</b>	68	7	10.3%	0.2%	89	8	9.0%	0.2%	65	3	4.6%	0.2%	39	3	7.7%	0.2%
<b>Total</b>	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 0.2%. 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. The process of regrading answer booklets resulted in 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>
<b>Phase 1</b>	8,546	51,276	7,943	47,658	4,645	27,870	4,577	27,462
<b>Phase 2</b>	1,008	6,048	937	5,622	644	3,864	695	4,170
<b>Phase 3</b>	68	70	89	92	65	68	39	42
<b>Total</b>	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  $\chi^2$  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						
	<u>2017</u>		<u>2018</u>		<u>Average</u>		<u>Diff</u>	<u>2018</u>		<u>2019</u>		<u>Average</u>		<u>Diff</u>
Group	<u>All</u> <u>Applicants</u>	<u>Applicants</u> <u>In Regrd</u>	<u>All</u> <u>Applicants</u>	<u>In Regrd</u>	<u>All</u> <u>Applicants</u>	<u>In Regrd</u>		<u>All</u> <u>Applicants</u>	<u>In Regrd</u>	<u>All</u> <u>Applicants</u>	<u>In Regrd</u>	<u>All</u> <u>Applicants</u>	<u>In Regrd</u>	
<u>Race</u>														
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%
<u>Sex</u>														
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%
<u>ABA?</u>														
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%
<u>CA ABA Tier</u>														
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.  $\chi^2$  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 direction<sup>7</sup>. 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) found in the various school groups may be a function of the

---

<sup>7</sup> 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  $\chi^2$  comparison tests were statistically significant at  $\alpha=.01$  level.

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 (SEM<sup>8</sup>). 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 granular look within the regrade range may provide further insights into making adjustments and refinements to the process.

---

<sup>8</sup> 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.

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  $\chi^2$  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  $\chi^2$  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)		
	<u>Test- Takers</u>	<u>Answer Boooks</u>	<u>% of Regrade</u>	<u>Test- Takers</u>	<u>Answer Boooks</u>	<u>% of Regrade</u>	<u>Test- Takers</u>	<u>Answer Boooks</u>	<u>% of Regrade</u>	<u>Test- Takers</u>	<u>Answer Boooks</u>	<u>% of Regrade</u>
<b><u>Regrade Band</u></b>												
<b>1435-1439</b>	107	642	10.6%	91	546	9.7%	40	240	6.2%	68	408	9.8%
<b>1430-1434</b>	100	600	9.9%	99	594	10.6%	80	480	12.4%	72	432	10.4%
<b>1425-1429</b>	104	624	10.3%	96	576	10.2%	63	378	9.8%	73	438	10.5%
<b>1420-1424</b>	115	690	11.4%	101	606	10.8%	59	354	9.2%	78	468	11.2%
<b>1415-1419</b>	95	570	9.4%	86	516	9.2%	72	432	11.2%	70	420	10.1%
<b>1410-1414</b>	121	726	12.0%	71	426	7.6%	66	396	10.3%	65	390	9.4%
<b>1405-1409</b>	97	582	9.6%	119	714	12.7%	47	282	7.3%	75	450	10.8%
<b>1400-1404</b>	80	480	7.9%	100	600	10.7%	60	360	9.3%	61	366	8.8%
<b>1395-1399</b>	103	618	10.2%	79	474	8.4%	94	564	14.6%	66	396	9.5%
<b>1390-1394</b>	86	516	8.5%	95	570	10.1%	62	372	9.6%	67	402	9.6%
<b><u>Regrade Total</u></b>	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<sup>9</sup> 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**

<sup>9</sup> 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**

	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)			
Regrade Band	<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%
<b>Regrade Total</b>	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 bands (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 0.1% 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.*

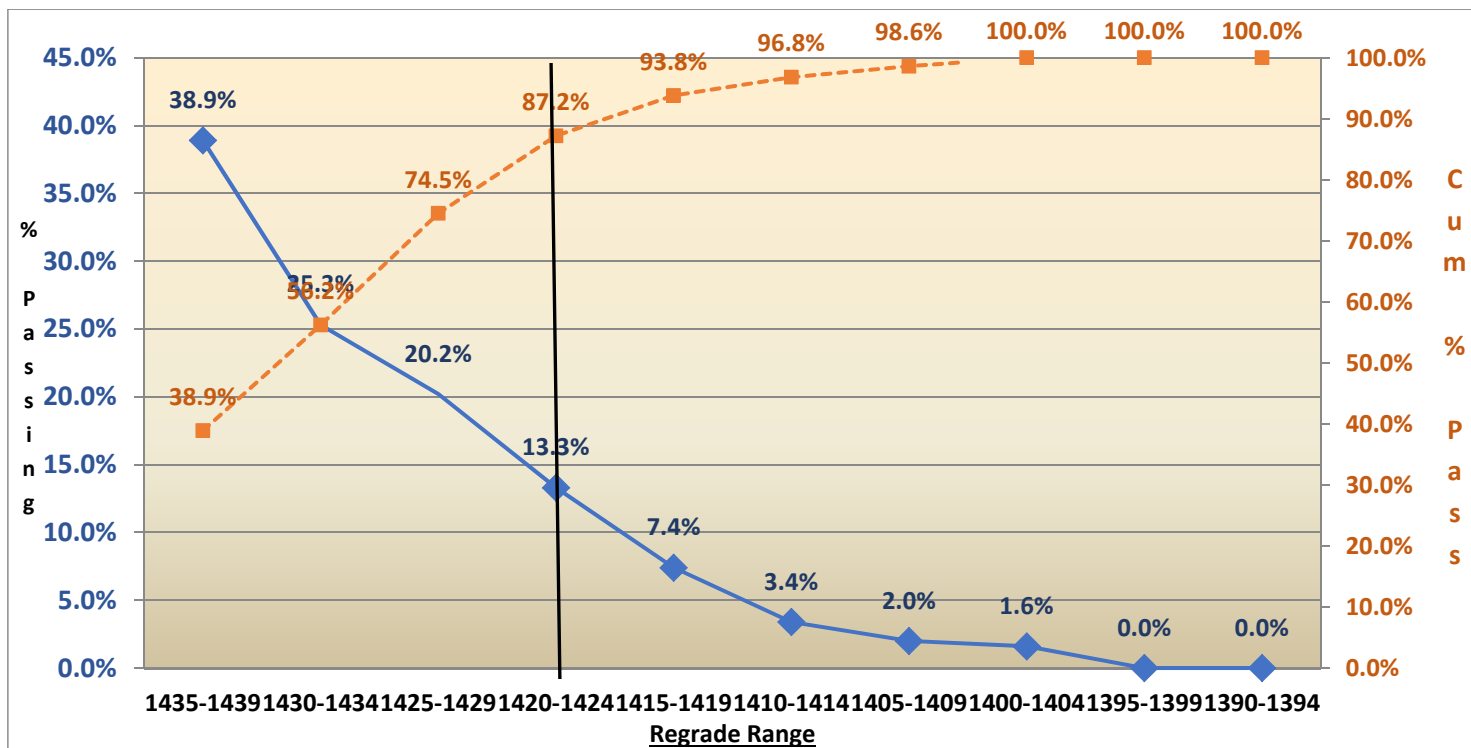
As demonstrated earlier, the impact of the regrade is relatively small, with about 3% of the passing rate in July and between 5% and 6% of the passing in February occurring during the process. Overall, the data in Table 5 suggests that the majority of that impact of regrade tended to occur for examinees 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 phase<sup>10</sup> and then compared the results. The resulting reliabilities are presented in Table 6 on page 23.

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

---

<sup>10</sup> Cronbach Alpha was used to estimate the reliability at each phase.



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>			
<b>2017</b>	.73	.73	.73
<b>2018</b>	.76	.76	.76
<b>Ave.</b>	.75	.75	.75
<u>February</u>			
<b>2018</b>	.70	.71	.71
<b>2019</b>	.70	.71	.71
<b>Ave.</b>	.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 format<sup>11</sup>. 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.<sup>12</sup> 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**

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

The results from Table 7 suggest that introduction of a 12<sup>th</sup> 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 (i.e., the questions were different on each examination, possibly effecting the

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

<sup>12</sup> 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.

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  $\chi^2$  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, respectively<sup>13</sup>. 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.<sup>14</sup>

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

---

<sup>13</sup> The estimates were inclusive of attorneys sitting for the written test only (1-Day Attorneys' Examination). They comprised about 13% of the total number of examinees in regrade.

<sup>14</sup> We further adjusted calculations based on the knowledge that our analysis sample excluded examinees taking the Attorneys' 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 0.5% 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 grading<sup>15</sup>.

The results of these estimations are presented in Table 8. What we observe from the calculation in Table 8 is that elimination of the regrade process would have resulted in fairly consistent decrease in the overall passing rates across the administrations of the CBX, varying by between 1.3% to 2.0%. 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

---

<sup>15</sup> 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.

schools would have been adversely affected. This latter finding was not consistent across all administrations.

**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>Days</u>
	<u>Rate</u>	<u>Decisions</u>				<u>Reliability</u>	<u>in Regrade</u>	<u>Books</u>	<u>Saved**</u>
								<u>Read</u>	
<b>Feb. 2018</b>	-1.3%	98.6%	No	No	Yes	.01	709	3,932	14
<b>Feb. 2019</b>	-2.0%	98.0%	No	No	No	.01	734	4,212	14
<b>Jul. 2017</b>	-1.6%	98.5%	No	No	No	.00	1,076	6,118	21
<b>Jul. 2018</b>	-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 (Attorneys') examination

Overall reliability of the examination was estimated to remain about the same (differing by no more than 0.01), 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 3<sup>rd</sup> Grading Phase-Resolution.** Presently, an entire week is allocated to the 3<sup>rd</sup> 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>				
<b>Feb. 2018</b>	-0.1%	99.9%	No	No	No	.00	65	68	7
<b>Feb. 2019</b>	-0.1%	99.9%	No	No	No	.00	39	42	7
<b>Jul. 2017</b>	-0.1%	99.9%	No	No	No	.00	68	70	7
<b>Jul. 2018</b>	-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 (Attorneys') examination

On each administration, the Resolution grading process serves to identify a small handful of applicants on each examination that could 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 0.1%. 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' 1<sup>st</sup> phase scores decreased, the chances of their passing decreased as well, to the point where no one passed with a 1<sup>st</sup> 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 2<sup>nd</sup> Phase regrade band, we evaluated a model where the band would be cut in half, i.e., from 1415 to 1439.99 instead of from 1390 to 1439.99. Under this model, examinees whose 1<sup>st</sup> Phase scores were between 1390 and 1414.99 would no longer enter regrade, but would be considered a “fail” at the end of the 1<sup>st</sup> Phase and the 3<sup>rd</sup> 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</u>	<u>Reduction</u>	<u>Overall Days Saved**</u>
			<u>Gender</u>	<u>Race</u>	<u>School *</u>		<u>No Longer in Regrade **</u>	<u>in Answer Books Read **</u>	
<b>Feb. 2018</b>	-0.1%	99.9%	No	No	No	.01	446	2,675	4.1
<b>Feb. 2019</b>	-0.1%	99.9%	No	No	No	.01	453	2,715	4.1
<b>Jul. 2017</b>	-0.1%	99.9%	No	No	No	.00	660	3,959	8.2
<b>Jul. 2018</b>	-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 (Attorneys') 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) 0.1% 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 administrations<sup>16</sup>. None of the examinees who would have failed under the modified process, but originally passed during regrade, entered the Resolution (Phase 3) process.

**Model 4: Maintaining Regrade Process-Increase the “Grading Rate”.** A 4<sup>th</sup> 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 administrations<sup>17</sup>. 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

---

<sup>16</sup> The 1-Day Attorneys’ Exam accounted on average for an estimated additional 14% examinees in regrade in February and 11% in July.

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

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 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 format<sup>18</sup> 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>	<u>Act.</u> Days	<u>Est.</u> Days	<u>Diff.</u>	<u>Act.</u> Days	<u>Est.</u> Days	<u>Diff.</u>	<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
<b>Total</b>	51	32	-19	51	31	-20	37	21	-16	37	20	-17

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

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 assumptions that the graders could each read and score 30 answers per day<sup>19</sup>. 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 examinee’s 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

---

<sup>19</sup> 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 Attorneys’ and GBX exams are included.

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 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		
	Act. Days	Est. Days	Diff.	Act. Days	Est. Days	Diff.	Act. Days	Est. Days	Diff.	Act. Days	Est. Days	Diff.
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 (14 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 grading 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**

<b><u>Modification</u></b>	<b><u>Difference in Pass</u></b>		<b><u>Consistency in</u></b>		<b><u>Differential</u></b>		<b><u>Written Section</u></b>	
	<b><u>Rate</u></b>		<b><u>Pass/Fail</u></b>		<b><u>Subgroup Impact*</u></b>		<b><u>Reliability Change</u></b>	
	<b><u>February</u></b>	<b><u>July</u></b>	<b><u>February</u></b>	<b><u>July</u></b>	<b><u>February</u></b>	<b><u>July</u></b>	<b><u>February</u></b>	<b><u>July</u></b>
<b>A. Structural</b>								
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 0.1% (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**

<b>Modification</b>	<b><u>Regrade Only</u></b>		<b><u>All Grading Phases</u></b>	
	<b><u>February</u></b>	<b><u>July</u></b>	<b><u>February</u></b>	<b><u>July</u></b>
<b>A. Structural</b>				
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>B. Graders</b>				
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., 2<sup>nd</sup> and 3<sup>rd</sup> 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 1<sup>st</sup> 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 2<sup>nd</sup> 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 Resolution process (Phase 3) alone was estimated to have the smallest estimated impact on passing rates (about 0.1%) 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 speed up results reporting by seven days.
- Model 3. A 3<sup>rd</sup> 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 0.1%, achieve a 99.9% consistency rate in pass/fail decisions (both identical to the estimates of eliminating Resolution grading), 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 5). 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. Since any modifications to the current phased grading system itself may be negatively perceived by the public, the State Bar might consider giving priority to system-based improvements to speed up release of grades.

As with other simulation type studies (e.g., those that pre-modeled the 2-Day examination format), the “et ceteris paribus” (i.e., all things held equal) assumption was in play during this study. 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 current high levels of examination reliability would remain intact if the pace of grading is increased. Both obtaining feedback from the grading teams and their leadership, as well as planning for an impact assessment would be essential before any changes are implemented.

## **References**

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