

**Profiling the Immediate and
Longer-Term Outcomes of the
California First-Year Law Students' Examination**

**An investigation of examinee performance and
potential opportunities for program changes**

A report prepared for the State Bar of California

By

Roger Bolus, Ph.D.

Research Solutions Group

April 26, 2022

Updated September 13, 2023

EXECUTIVE SUMMARY

California's First-Year Law Students' Examination (FYLTX) program is primarily intended to provide students from California's unaccredited law schools with an assessment of whether they are making progress in their legal education with the end goal of graduating from law school, passing the California Bar Examination (GBX) and eventually entering the practice of law. The program has existed in its current form for over 23 years and since the year 2000 has tested 24,075 examinees.

This investigation was designed to provide a broad historical perspective on the composition of the student population that has sat for the exam, their performance, both in terms of test scores and passage rates, and their eventual experience on the GBX. The study was also designed to offer insights into structural aspects of the examination including its composition, scoring, and decision rules regarding student passing. The analyses were conducted with the objective of possibly reducing the testing burden on students, improving the efficiency of delivering the exam while simultaneously maintaining the program's psychometric excellence.

Through a review of examination results and demographics of 10,340 unique students who sat for the FYLTX for the first time between June 2000 and October 2016, the study found the following:

- ***Whites comprised 54% of FYLTX test takers, with Asians, Hispanics and Blacks comprising between 10% and 11%, respectively.*** The relative proportion of White test takers has fallen steadily over time to 49% in recent years, as has the proportion of males (53%). Students from Unaccredited Law Schools historically make up 85% of the test takers.
- ***Only 27% of the first-time test takers pass the FYLTX on their initial attempt.*** About 2/3 of those who fail will make further attempts to pass. Most students (76%) will make only a few additional tries, but a sizable percentage (24%) will make 4 or more attempts.
- Overall, 74% of those who did make a subsequent attempt at passing the FYLTX showed some improvement in their scores by their final attempt (50% will improve by almost 40 points). ***Only 36% of those who initially failed however eventually passed, resulting in a net overall FYLTX pass rate of 44%.***
- ***The likelihood of passing the FYLTX was strongly related to students' initial performance.*** For example, students who initially scored relatively close to the 560-point passing standard had a 76% likelihood of eventually passing, while those initially scoring less than 400 had less than a 1 in 10 chance of passing (despite showing similar amounts of improvements in their scores).
- After controlling for initial FYLTX performance, no differences were observed between males and females in the likelihood of eventually passing. ***Blacks were somewhat less likely to pass than***

the other racial/ethnic groups; however, there was insufficient data to determine the root cause of that difference.

- ***Only 35% (N=3,616) of the initial sample of 10,340 FYLSX takers sat for the GBX.*** These included 90.3% who had passed the FYLSX and a small sample of 9.7% who failed, but found alternative means of qualifying. The demographic composition of those who eventually took the GBX was fairly similar to the composition of the original group, with no group differing by more than 5%.
- ***29% of the 3,616 students passed the GBX on their initial try¹, far below the 62% average GBX passage rate across almost all 150,000 first-time GBX takers over the same period.*** Average GBX test scores of the FYLSX sample were about 110 scale score points lower than all GBX takers.
- ***83% (N=2,310) of those FYLSX applicants who failed on their initial GBX attempt made subsequent attempts to pass, a rate almost double that of the initial failing FYLSX takers who reattempted the FYLSX.***
- ***80% of GBX repeaters experienced some improvement in their GBX performance, improving by an average of 77 scale score points.*** The final GBX passage rates after all attempts for the sample of 3,616 students was 57% (1,440) and 62% (1,390); these rates were almost double the passage rates on the students' initial attempt.
- ***Across all 10,340 first-time FYLSX in the analysis sample, only 19% eventually passed the GBX.*** The major percentage (53%) of the sample failed the FYLSX and never attempted to take the GBX. There were two equally sized groups (13% each) that passed the FYLSX; one group opting not to take the GBX at all, while the other took the GBX and failed. Both groups had comparable FYLSX scores. No data are available on the factors that contributed to these examinees' decisions.
- ***Logistic regression models revealed that FYLSX MC and Total scores were highly predictive of GBX outcomes, which speaks to the strong predictive validity and value of the FYLSX as a signaling device for students.*** For example, students scoring at the lower ends of the FYLSX passing scoring range (541-560) had a 41% predicted (39% actual) GBX passage rate, while those in the upper range (681-700) had a 92% predicted (93%) actual passage rate.

Aspects of these findings also pointed to possible modifications changes that could be made to the structure and scoring of the FYLSX; changes that could be implemented, without negatively impacting the quality of the test or student outcomes. To examine potential alternatives, the following additional analyses were conducted:

- First, the relatively low reliability of the Essay section, in tandem with the high reliability of the Multiple-Choice section, ***suggested that the overall reliability of the FYLSX and its predictive validity could be improved by either more heavily weighting the MC section (relative to the***

¹ Analyses of GBX passage was done using both the 1440 standard as well as an "adjusted" 1390 standard.

Essay section) or possibly eliminating the Essay section completely. The latter finding, also pointed out in previous reports, would reduce testing burden on students while offering logistical opportunities to the State Bar. Further analyses also indicated that either of these **changes could be made without adverse impact on any particular group of students.**

- Because the FYLSX passage rates were found to be consistently high throughout the 20-point Reappraisal range (73% to 100%), analyses were conducted simulating the impact of eliminating Reappraisal completely. **Two alternatives were evaluated: (a) simply dropping the passing score to 540, or (b) maintain the 560 passing score, thereby failing those who passed as the result of Reappraisal.** Option (a) was found to have the smallest impact, affecting the final outcomes of only .8% of the 10,341 students.
- Finally, based upon the strength of the logistic regressions mentioned above, additional analyses were conducted **to estimate the impact of lowering the passing standard below the current 560 level.** The likelihood of passing the GBX at differing FYLSX score ranges was estimated, along with the incremental numbers of students that might have eventually passed the GBX and the resulting passage rates in the total sample. The analyses revealed that while the absolute number estimated to pass the GBX would gradually increase, the overall passage rates would decrease. **Pro-rating the results to a per-administration basis indicated that, on average, lowering the FYLSX passing standard would affect the GBX outcomes of only 1 to 2 students per 10 to 20-point drop in the FYLSX passing standard.** A strong cautionary statement was made regarding this modeling exercise.

We believe that these findings should provide policy makers with important foundational information as they consider the future directions of the FYLSX program.

I. BACKGROUND

A. Purpose of the Examination.

California's First-Year Law Students' Examination (FYLSX) is primarily intended to provide students from California's unaccredited law schools with an assessment of whether they are making progress in their legal education with the end intent of graduating from law school, passing the California Bar Examination (GBX) and entering the practice of law. Students from these schools must pass the FYLSX to receive credit for their initial year of legal education. The FYLSX is also used for other purposes. For example, students who have been disqualified from ABA schools for academic or other reasons can be required to pass the FYLSX before they can be readmitted. Law Students who have not completed a formal college education, or have arranged to receive their legal training through judges' chambers (so-called "Special" examinees), are also required to take and pass the FYLSX before they can continue on in their training. And yet another small group take the exam simply as a means of practice.

B. Test Structure.

The FYLSX has been administered for over 40 years. While several modifications were made to the examination in the early years, it has been administered, graded and scored in its current form since 1999. The FYLSX, administered twice a year in June and October, consists of two sections administered in a single day. The first section is a written, 4-hour, test consisting of four essay questions, each graded on a 40 to 100-point scale (in 5-point increments). Questions are developed by a team of lawyers managed by the State Bar of California. A new set of questions are developed for each administration. The second FYLSX section is a 100-item multiple-choice (MC) test. There are currently five forms of the MC test with different sets of items (some of which overlap). Each form has two layouts with items located in different positions to minimize the possibility of cheating. The current forms were developed in 1998-1999.

C. Test Scoring.

The raw score on the MC test is simply the number of items answered correctly. That raw score is converted to a MC Scale Score through a process known as "equating". The purpose of the equating process is to control for differences in test difficulty between forms and over time, through the use of commonly appearing items on the various forms. The theoretical range of scale scores is 0 to 800 points. The raw score on the essay section is the sum of the scores earned on each of the four questions. To control for any differences in question difficulty or grader leniency, the raw scores are subsequently

placed on the same scale of measurement as the scaled MC test creating an Essay Scale Score. A Total Scale Score is calculated by summing the MSC and ESC, which effectively gives equal weight to both sections. Examinees achieving a Total Scale Score of 560 or higher pass the FYLSX. Those scoring below 540 will fail. If their Total Scale Score falls between 540 and 559.999, an examinee will have their answers reappraised by the supervisor of the grading team. Based on this review, scores are not changed, but a final pass or fail decision is made². The same test scoring process and pass/fail/reappraisal standards have been in place since 1999.

D. Previous Statistical Reports.

There have been several statistical reports on the FYLSX. After each administration, a brief technical report is prepared documenting the psychometric properties of the test, along with salient descriptive test statistics. The focus of these reports is designed to meet technical testing standards and are limited to the current administration only. Three reports (Klein 1982 and 1999, Klein & Bolus 2010) have looked at the FYLSX more broadly (i.e., over multiple administrations). The purpose of these reports was primarily to study the relationships between FYLSX performance and GBX performance with an eye towards considering possible changes in the structure and scoring of the FYLSX itself.

These reports are now quite dated. Between the initial report in 1982 and the 2010 report, many changes have occurred including the structure of both the FYLSX and the GBX, the scoring methods for those examinations, and the nature of the examinees who sat for those examinations. In addition, the landscape of legal education has changed in California, particularly in the accreditation of law schools. The same can be said for the data available to those historical reports and the present one. As a result, while those previous reports can serve as a reference point for policy makers, their value is limited. As the State Bar of California begins to consider modifications to the FYLSX testing program, it is clearly time to update and expand the findings from those previous investigations.

² Thus, it is possible to have a student pass while his registered FYLSX score falls between 540 and 559.999.

II. PURPOSE OF THE STUDY

The Standards for Educational and Psychological Testing, the de facto guide for professional testing, recommends that testing programs such as the FYLSX be reviewed every seven to 10 years to assure relevance, validity and reliability. Given that it had been 12 years since the last review of the FYLSX, the State Bar of California (“The State Bar”) and the Committee of Bar Examiners (“The Committee”) commissioned an investigation that would replicate the previous studies but which would now focus on more recent FYLSX data along with expanded information on examinees which was not available at the time of the earlier studies. By investigating both the near- and longer-term outcomes of examinees who have taken the FYLSX, and evaluating changes that have occurred over time, results from this study are intended to aid the State Bar policy makers in making potential modifications to the FYLSX content, grading procedures and scoring, and possibly, passing standards.

A. Policy Issues.

This study investigated the following issues regarding the FYLSX:

1. The viability of the current FYLSX testing structure, i.e., a multiple-choice component and a written component.
2. The continuing requirement for a reappraisal phase during the grading and scoring process.
3. The appropriateness of the scoring methods, including the current component weighting rules; i.e., each component gets 50% weighting.
4. The appropriateness of the current FYLSX passing standard of 560 scale points, which was established approximately 25 years ago.
5. The possibility of providing additional statistical results to the public that could aid students in their decision-making process to continue law school, and eventually take the GBX.

B. Analysis Questions.

The following questions regarding examinee performance on the FYLSX, and subsequently on the GBX, structured the statistical analyses for this study.

1. How have examinees performed on the FYLSX, including:
 - To what degree has performance changed over time?
 - To what degree has performance differed by the type of law school and demographic composition of the population of test takers?
 - How did the reappraisal process impact the passage rates?
2. What was the subsequent FYLSX experience for examinees who failed the FYLSX on their first attempt, including:
 - How many additional attempts would they make to pass the FYLSX?
 - What were their eventual outcomes?
 - How much improvement was made on their subsequent attempts?
 - To what degree were subsequent outcomes related to student demographics or the law school they attended?
3. What was the GBX experience for examinees who took the FYLSX, including:
 - How many students eventually took the GBX?
 - How many attempts did it take them to pass the GBX?
 - To what degree was performance on the FYLSX related to performance on the GBX?
 - To what degree was examinees' performance influenced by the type of law school attended or demographic characteristics of the test takers?
 - Was it possible to accurately predict who would eventually pass the GBX based on their FYLSX performance? If so, did student demographics and/or type of law school affect this prediction?
 - How might these results have changed if the historic GBX passing standards of 1440 were reset to the current 1390 level?

Answers to these questions are intended to (a) provide an evidence-based foundation for considering the policy issues identified above, and (b) serve as a baseline for assessing any changes to the program.

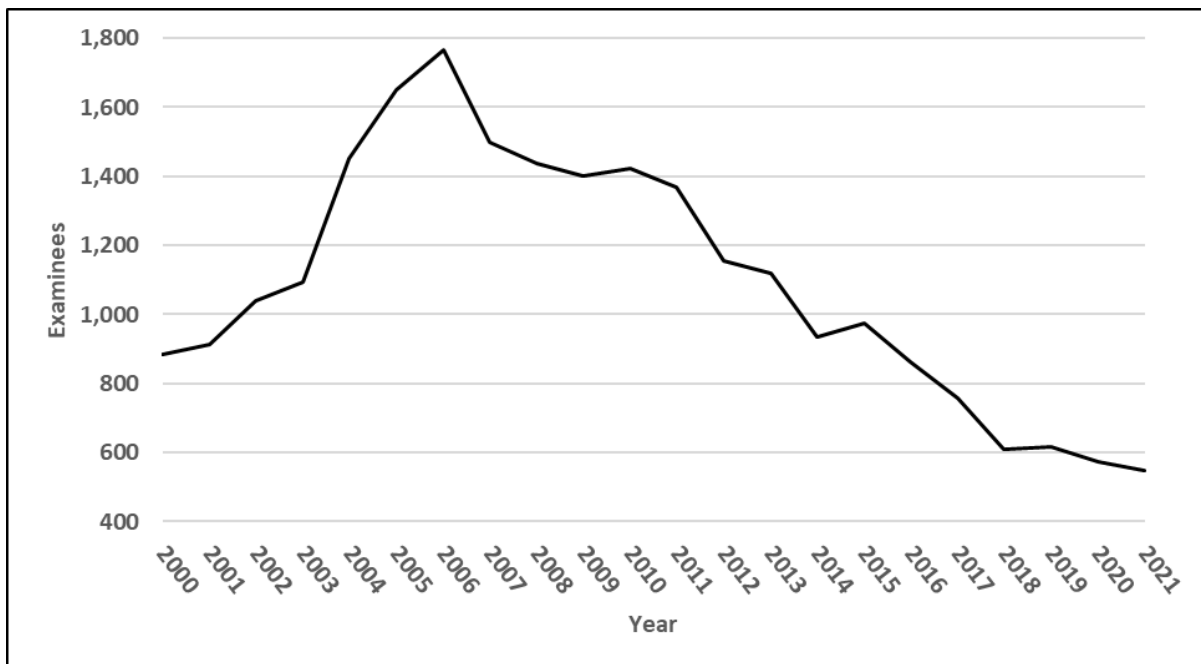
III. STUDY DATA

A. All FYLSX Takers.

As detailed below, the current study expands beyond prior analyses by extending the number of examinations (and therefore the number of students) that are included and incorporating, where possible, additional demographic characteristics of examinees³.

Between June 2000 and October 2021, there have been 44 administrations of the FYLSX. During that period, 24,075 examinees sat for the examinations; 54% during the June testing and 46% during the October testing. The number of test takers per year (June plus October) rose between 2000 and 2006, and then began a steady decline through 2021, when the number of examinees fell to a low of 541 from a high of almost 1,800 in 2006. Figure 1 provides a graphic illustration of the annual number of test takers.

Figure 1
Annual Count of FYLSX Test Takers
2000-2021



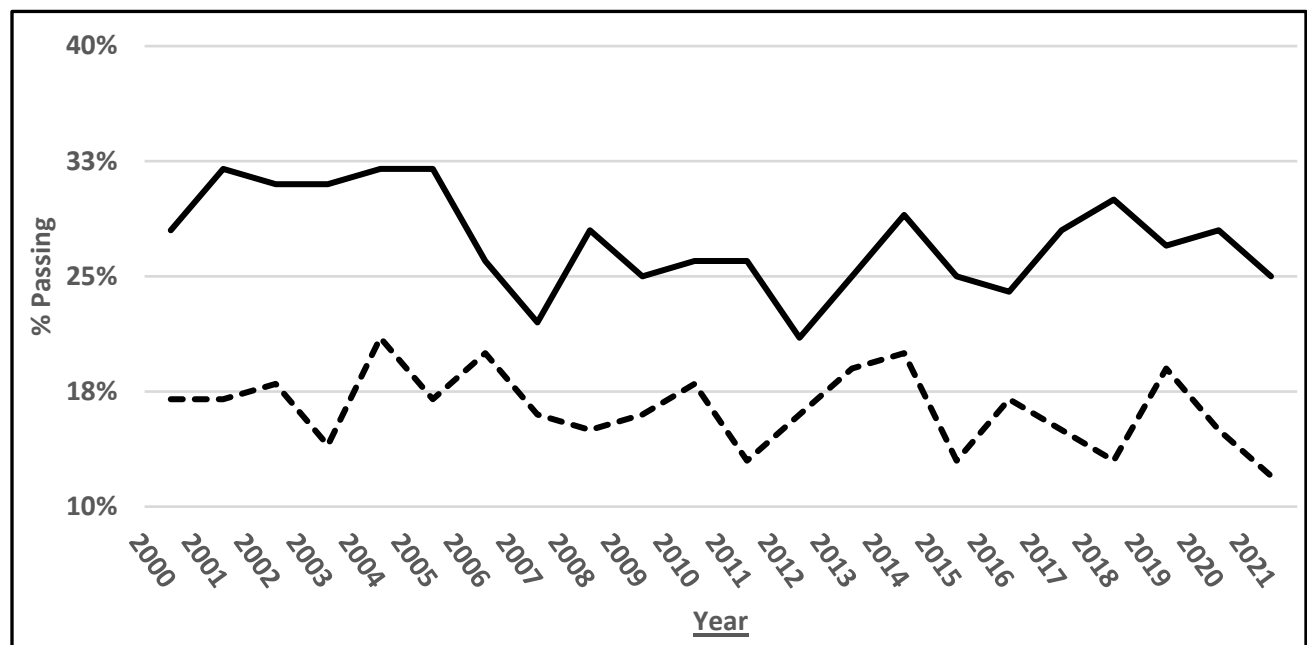
The examinee pool across this period was comprised of 11,734 (48.7%) first-time takers and 12,341 repeaters (51.3%). Given that the primary target for the FYLSX is California unaccredited law schools, it's

³ Note that we refer to "students," "examinees" and "test takers" interchangeably throughout this report.

not surprising that over the 22 years, students from these schools comprised the majority of test takers (84.5%); (39.7% from Distance Learning institutions, 24.0% from Correspondence Schools and 20.8% from Fixed-Facility schools). The balance of students came from California Accredited schools (7.3%), 5.0% from students who were disqualified from ABA approved law schools, 2.3% from students who received instruction in Judge's Chambers and less than 1% who came from other types of legal education or had no school assigned to their electronic record.

While the passing rate on the FYLSX has been historically quite low, 21.95% overall, it has vacillated over the years ranging from a high of 27% to a low of 17% in 2021. The overall passing rates of 1st-time takers and repeaters differed by 10% over the 22 years (27% vs 17%); however, the size of that difference has been as large as 17% in 2018 and as small as 5% in 2012. Figure 2 illustrates these differences over time.

Figure 2
FYLSX Passing Rates by Year for 1st-Time Takers and Repeaters
2000 through 2021



B. Study-Specific Sample.

Because of the substantial differences in passing rates between 1st-timers and repeaters, we reasoned that these two groups would be inherently different. To establish a cohort that would be most homogeneous relative to their FYLSX experience, we first created a sample with 1st-time takers only. Since

the majority of research questions focused on *eventual* FYLSX performance outcomes, we reasoned that a longitudinal design (i.e., tracking the outcomes of individual test takers over time) would be more appropriate than a cross-sectional one, allowing each test taker to serve as their own “control”.

For each examination-specific cohort of first-time examinees (e.g., June 2000), we identified a fixed follow-up period consisting of the next 10 administrations of the FYLSX (e.g., through October 2005 for those first-time examinees from June 2000). A search of data for the follow-up period was conducted to determine if failing examinees from the initial cohort made subsequent attempts, and if so, how many. By maintaining a constant follow-up period, no one examinee would benefit by having more opportunities to pass.⁴ Employing this logic, it was possible to develop 34 separate cohorts of first-time FYLSX takers beginning in June 2000 and extending through October 2016.⁵ Ending the FYLSX study cohort at October 2016 provided a sufficient follow-up period to track these examinees through subsequent attempts to pass the GBX. The resulting sample size was 10,340 1st-time FYLSX takers.

Table 1 on the next page provides a description of the composition of the study sample in terms of five examinee characteristics: race/ethnicity, gender, school (or program) attended, number of FYLSX attempts, and examinee classification regarding their reason for taking the exam (i.e., as a student who was “Disqualified” from an ABA or California Accredited school, as a “Special” student who was required to take the examination because of their study in Judge’s Chambers, or because they attended Unaccredited institutions (“Regular”). To investigate whether any temporal trends emerged over the course of the full 17-year study period, we stratified the data into three separate 4-year segments, and a final 5-year segment (2012-2016).

The relative demographic composition and source of legal education is presented for each of these four time periods (and across the entire study period) in Table 1. The table reports the relative percentage (out of 100%) that each subgroup of examinees represented out of the total number of examinees *within each time period*. The final column in the table presents the percentages expressed across the full 17-year period.

With respect to Race/Ethnicity, Table 1 shows that since 2007 the percentage of Whites sitting for the FYLSX has declined from greater than a half of test-takers (57%) to just under a half (49%). This drop has been accompanied by a corresponding increase in the relative proportion of those of Hispanic Origin (from

⁴ Analyses showed that 90% of examinees took the FYLSX only 3 times, 95% took the exam no more than 5 times, and 99% made 8 or less attempts.

⁵ This extended the time period covered in the Klein & Bolus (2010) study by an additional 10 years.

9% to 15%). ⁶ Corresponding to recent trends on the GBX, females have made up an increasing greater proportion of test takers over the 16-year

Table 1
Relative Composition of
FYLSX 1st-Time Test Takers by Year

| | FYLY Examination Period | | | | |
|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|-----------------------------|
| Category | 2000-2003 (N=2,168) | 2004-2007 (N=3,223) | 2008-2011 (N=2,686) | 2012-2016 (N=2,263) | Total (N=10,340) |
| <u>Race/Ethnicity</u> | | | | | |
| Asian | 9% | 10% | 11% | 10% | 10% |
| Hispanic | 9% | 10% | 12% | 15% | 11% |
| Black | 11% | 10% | 13% | 12% | 11% |
| White | 57% | 57% | 51% | 49% | 54% |
| Other | 5% | 6% | 8% | 8% | 7% |
| Missing | 9% | 7% | 6% | 6% | 7% |
| <u>Gender</u> | | | | | |
| Male | 58% | 59% | 54% | 53% | 56% |
| Female | 36% | 39% | 42% | 44% | 40% |
| Missing | 6% | 2% | 4% | 3% | 4% |
| <u>School</u> | | | | | |
| ABA | 6% | 6% | 6% | 5% | 6% |
| Accredited | 8% | 6% | 6% | 5% | 6% |
| Unaccredited | 82% | 86% | 86% | 87% | 85% |
| <i>Correspondence</i> | 24% | 19% | 24% | 27% | 23% |
| <i>Distance Learning</i> | 36% | 51% | 46% | 42% | 45% |
| <i>Fixed-Facility</i> | 22% | 16% | 16% | 18% | 18% |
| Other | 3% | 2% | 2% | 4% | 2% |
| <u>No. of Attempts</u> | | | | | |
| 1 Attempt | 56% | 52% | 53% | 51% | 53% |
| 2 Attempts | 22% | 24% | 23% | 25% | 24% |
| 3 Attempts | 12% | 12% | 13% | 13% | 12% |
| 4 Attempts | 4% | 6% | 5% | 5% | 5% |
| 5 or more | 6% | 6% | 6% | 6% | 6% |
| <u>Type of Examinee</u> | | | | | |
| Regular | 87% | 90% | 90% | 93% | 90% |
| Disqualified | 11% | 9% | 9% | 6% | 9% |
| Special | 2% | 1% | 1% | 1% | 1% |

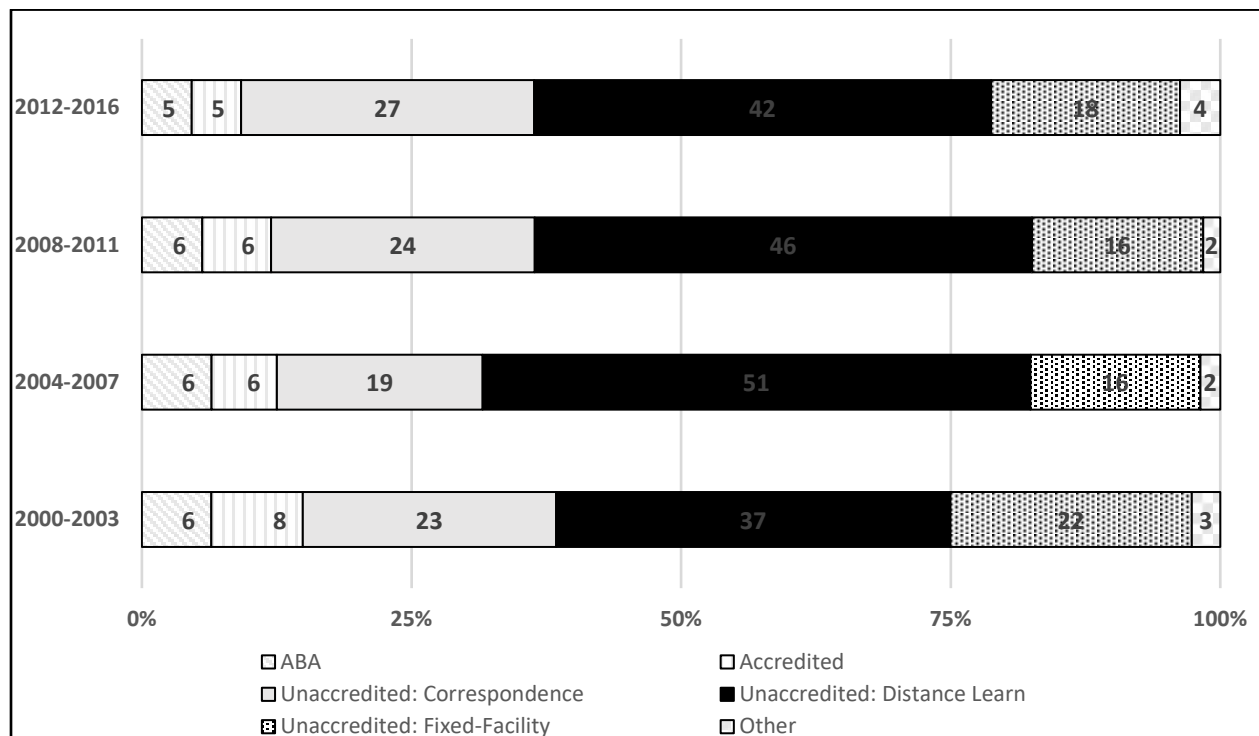
period, increasing from 36% to 44%. The relative proportion of test takers making their initial attempt while attending an Unaccredited school has increased slightly over the 16-year period (from 82% to

⁶ Note the trend has continued in the most recent 5 years where Whites made up 47% and Hispanics 20%.

87%). Due in part to the opening and closing of schools over the period, Table 1 shows that the relative percentage of students from the three types of unaccredited schools has seesawed. Across all periods, the majority students have attended Distance Learning programs (ranging between 36% and 51% of examinees). (The relative rates of the type of legal program that the FYLSX takers attended are also further illustrated in Figure 3.) Table 1 also illustrates that the number of attempts made by students to pass the FYLSX remained relatively stable over time. The percentages of students that made two or less attempts ranged between 77% and 78% of all takers, while those making 5 or more attempts remained consistent over the 16-year period. (6% of all test takers). Finally, we observed a slight downwards trend in the proportion of Disqualified students over the 16-year period, dropping from 11% in 2000 to 2003 down to 6% in 2012 to 2016. This drop corresponds to the decreases in the proportion of students from ABA and California Accredited law schools over the same time period.

Figure 3

**Relative Composition of FYLSX Examinees by Type of Legal Education
2000-2021**



With this description of the analysis sample as a backdrop, the balance of the report presents statistical evidence addressing each of the study research questions.

IV. FINDINGS

A. What was the initial and eventual performance of examinees taking the FYLSX?

1. Initial FYLSX Score Performance. Table 2 presents average initial performance scores on the two FYLSX sections and combined Total Scale Score by year and across all 16 years for all 10,340 first-time test takers, further broken down by demographics and type of school.

Table 2

**Initial FYLSX Total Scale Scores for
FYLSX 1st-Time Test Takers by Year**

| Category | Multiple-Choice | | | | | Essay | | | | | Total | | | | |
|----------------|-----------------|---------------|---------------|---------------|--------------|---------------|---------------|---------------|---------------|--------------|---------------|---------------|---------------|---------------|--------------|
| | 2000- 2003 | 2004- 2007 | 2008- 2011 | 2012- 2016 | All Years | 2000- 2003 | 2004- 2007 | 2008- 2011 | 2012- 2016 | All Years | 2000- 2003 | 2004- 2007 | 2008- 2011 | 2012- 2016 | All Years |
| Overall | 247 | 246 | 240 | 238 | 243 | 247 | 247 | 241 | 242 | 244 | 494 | 492 | 481 | 480 | 487 |
| Race | | | | | | | | | | | | | | | |
| Asian | 244 | 243 | 238 | 232 | 239 | 237 | 246 | 236 | 239 | 240 | 481 | 489 | 474 | 471 | 479 |
| Hispanic | 228 | 229 | 227 | 223 | 227 | 235 | 235 | 231 | 237 | 235 | 463 | 464 | 458 | 460 | 461 |
| Black | 217 | 214 | 218 | 216 | 216 | 224 | 220 | 223 | 221 | 222 | 441 | 435 | 442 | 437 | 439 |
| White | 260 | 258 | 254 | 253 | 256 | 256 | 255 | 251 | 251 | 254 | 517 | 513 | 505 | 504 | 510 |
| Other | 230 | 232 | 227 | 228 | 229 | 243 | 239 | 236 | 237 | 238 | 472 | 470 | 463 | 464 | 467 |
| Missing | 233 | 231 | 221 | 224 | 228 | 233 | 237 | 228 | 234 | 233 | 466 | 468 | 449 | 458 | 461 |
| Gender | | | | | | | | | | | | | | | |
| Male | 254 | 251 | 248 | 245 | 250 | 248 | 245 | 242 | 242 | 244 | 502 | 496 | 491 | 487 | 494 |
| Female | 240 | 239 | 233 | 231 | 235 | 247 | 250 | 241 | 244 | 245 | 487 | 489 | 473 | 475 | 481 |
| Missing | 229 | 208 | 214 | 220 | 219 | 226 | 226 | 225 | 229 | 226 | 455 | 434 | 439 | 448 | 445 |
| School | | | | | | | | | | | | | | | |
| ABA | 229 | 239 | 242 | 224 | 235 | 241 | 252 | 255 | 238 | 248 | 470 | 491 | 497 | 462 | 482 |
| Accredited | 229 | 224 | 217 | 213 | 222 | 235 | 236 | 235 | 225 | 233 | 464 | 460 | 452 | 438 | 456 |
| Unaccredited | 251 | 249 | 242 | 241 | 246 | 249 | 247 | 241 | 244 | 245 | 502 | 493 | 491 | 489 | 493 |
| Correspond. | 254 | 248 | 246 | 245 | 248 | 249 | 245 | 245 | 244 | 245 | 502 | 493 | 491 | 489 | 493 |
| Distance | 261 | 255 | 245 | 245 | 251 | 253 | 251 | 241 | 246 | 248 | 514 | 506 | 486 | 491 | 499 |
| Fixed | 233 | 229 | 228 | 223 | 229 | 241 | 237 | 233 | 238 | 237 | 474 | 466 | 461 | 461 | 466 |
| Other | 220 | 212 | 232 | 230 | 224 | 235 | 230 | 229 | 235 | 233 | 455 | 442 | 461 | 465 | 456 |
| Type | | | | | | | | | | | | | | | |
| Regular | 250 | 248 | 242 | 240 | 245 | 248 | 247 | 241 | 243 | 245 | 498 | 494 | 482 | 483 | 489 |
| Disqualified | 219 | 226 | 226 | 214 | 223 | 228 | 242 | 242 | 229 | 237 | 447 | 468 | 468 | 444 | 459 |
| Special | 273 | 270 | 269 | 240 | 264 | 272 | 263 | 278 | 236 | 262 | 544 | 533 | 548 | 476 | 526 |

As a means of aiding in the interpretation of differences between the group averages (exam period, law school, and demographics) means, we also developed a series of linear models to determine the “percentage of variation” (POV) between examinees scores that could be accounted for by knowledge of

their group membership. The “POV” could range anywhere from 0% to 100%⁷. The higher the percentage, the greater the influence that being a member of a specific category had on the outcome. For the purpose of this report, we consider POV values less than 5% should be considered as minimal practical impact.

Inspection of Table 2 reveals that across the 17-year period, the average Total Scale Score was 487, which was 73 points below the passing standard of 560. While the overall POV related to time was less than 1%, we did observe a slight average drop of 14 scale score points over the years

The findings were similar for the Multiple-Choice and Essay Scale scores. With respect to school type, no significant differences were found for Essay performance (POV=0%), MC performance (POV=4%), nor overall Total Scale scores (POV=3%). Similarly for Gender, no significant differences were observed for Essay performance (POV=.0%), MC performance, nor Total Scale scores (POV=4%).

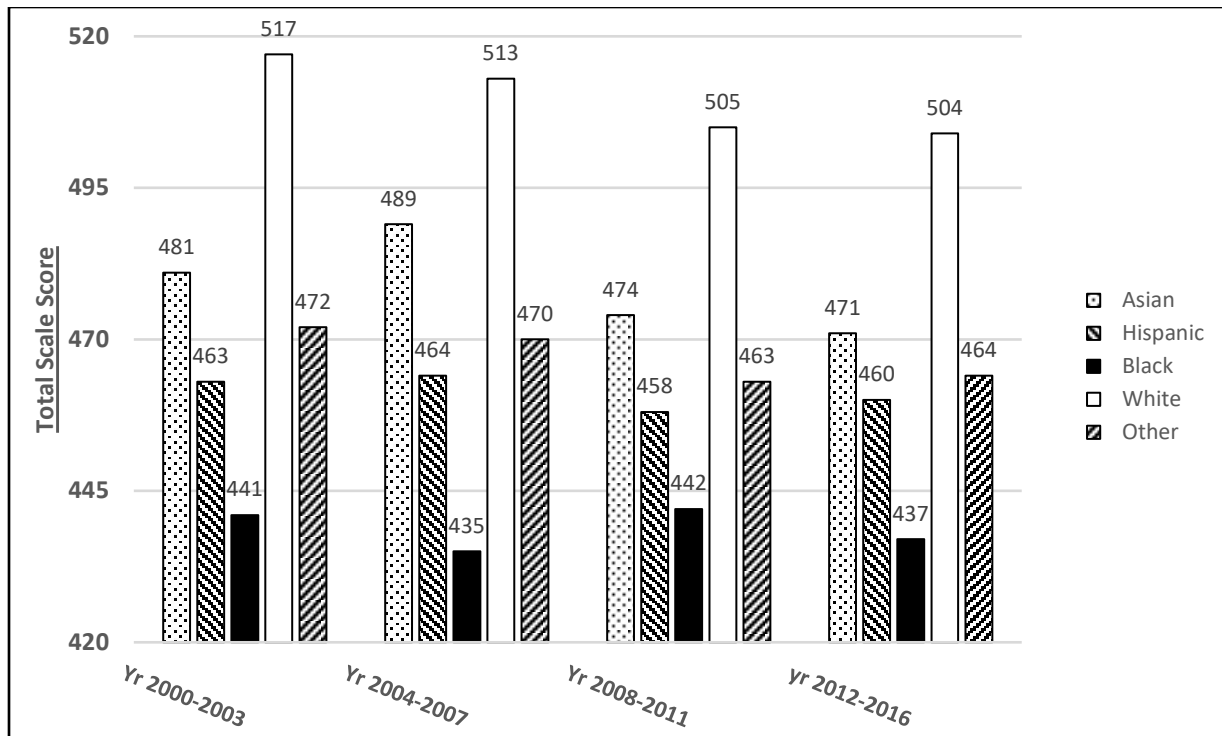
Of note however, was the relationship between Racial/Ethnic membership and initial FYLSX performance. For both the MC scores and overall Total Scale scores, the POV was 10%, while the POV for the Essay Scale scores was 5%. Figure 4 illustrates the mean Total Scale scores for each racial/ethnic group for each time period. Table 3 and Figure 4 indicate that White students had consistently scored higher on the FYLSX than minority students. Over the 16-year period, White students have scored 71 points higher than Black students (510 vs. 439), 49 points higher than Hispanics (510 vs. 461), and 31 points higher than Asians (510 vs. 479). We concluded that a small, but meaningful, difference in FYLSX performance could be attributed to race/ethnicity. *Of further interest is the finding that the size of the POV’s associated with the MC, Essay and Total Scores are almost identical to that observed in the GBX for both racial/ethnic groups and gender, differing by only 1% in POV.* Given the difference in time between administration of the FYLSX and the GBX, this finding might suggest that any differences that have been observed may be more of a function of the ability levels of the groups rather than inherent biases in the respective examinations.⁸

Figure 4

**Average Initial FYLSX Total Scale Score by Racial Ethnic Group
2000-2021**

⁷ As a point of reference for interpreting POV values, a previous study showed indicated that the POV for predicting GBX scores from law school GPA was 50%.

⁸ An additional set of models were run adding the interaction of “Time Period” to each of the law school and demographic groups. The purpose of these models was to assess whether any differences observed between the groups shifted with time. In all models, no statistically significant interactions were found. This suggested that performance differences between the various groups have remained consistent over time.



2. Initial FYLSX Passage Rates. Table 3 presents the initial FYLSX passage rates associated with the scores from Table 2. The overall initial passing rate across the 10,340 students taking the FYLSX for the first time was 27%. Along with the slight downward trend in FYLSX Total Scale scores came a slight, but steady, decrease in passage rates (from 30% in the beginning of the time period to 25% in the latter years). The magnitude of the differences in the average scores corresponded with the differences in passage rates. For example, the large POV for race was reflected in the differences in overall passage rate among whites (36%) vs. Asians (26%), Hispanics (18%) and Blacks (9%). Interestingly, the 5% difference between male and female passing rates are identical to the differences that have been observed on the GBX. Not unexpectedly, students from the Unaccredited schools, the primary target of the FYLSX, passed at slightly higher rates than those students from either ABA or California Accredited schools; students who were primarily taking the exam for poor law school performance or disciplinary reasons.

Table 3
Initial FYLSX Passage Rates
FYLSX 1st-Time Test Takers by Year
2000-2021

| | - | | | | |
|--------------------|------------------|------------------|------------------|------------------|------------------|
| Category | 2000-2003 | 2004-2007 | 2008-2011 | 2012-2016 | All Years |
| Overall | 30% | 28% | 26% | 25% | 27% |
| Race | | | | | |
| Asian | 27% | 27% | 27% | 20% | 26% |
| Hispanic | 21% | 17% | 19% | 18% | 18% |
| Black | 11% | 6% | 11% | 8% | 9% |
| White | 39% | 36% | 35% | 34% | 36% |
| Other | 21% | 20% | 17% | 23% | 20% |
| Missing | 18% | 22% | 10% | 13% | 16% |
| Gender | | | | | |
| Male | 33% | 30% | 30% | 27% | 30% |
| Female | 28% | 27% | 24% | 23% | 25% |
| Missing | 15% | 6% | 3% | 11% | 9% |
| School | | | | | |
| ABA | 16% | 27% | 25% | 13% | 21% |
| Accredited | 21% | 12% | 10% | 10% | 14% |
| Unaccredited | 33% | 30% | 28% | 26% | 29% |
| <i>Correspond.</i> | 38% | 31% | 33% | 29% | 32% |
| <i>Distance</i> | 37% | 33% | 27% | 27% | 31% |
| <i>Fixed</i> | 20% | 18% | 20% | 20% | 22% |
| Other | 19% | 20% | 21% | 25% | 22% |
| Type | | | | | |
| Regular | 32% | 29% | 27% | 26% | 28% |
| Disqualified | 11% | 17% | 16% | 12% | 14% |
| Special | 56% | 43% | 50% | 28% | 45% |

3. Eventual FYLSX Performance. Of the 10,340 examinees who sat for the FYLSX, 2,836 passed on their initial attempt (27.4%), while the remaining 7,504 (72.6%) failed. Of these 7,504, only about 2/3 (4,873; 47% of all 1st-time takers) chose to make further attempts to pass. Among those repeating the exam, half (2,433; 50.1%) made one additional attempt; another quarter (1,258; 25.9%) made two more attempts;

13% attempted three more 3 times; and the balance (24%) of the 4,873 examinees made anywhere from 4 to 10 tries at passing.

We then sought to determine what was going to happen to students after failure on their initial attempt and if their eventual outcomes could be predicted in some manner from their initial performance. Answering this question began by first gaining an understanding of *how well initial FYLSX scores were related to eventual scores (i.e., establishing their correlations)*, and then determining the *size of the differences after subsequent attempts*.

Table 4 illustrates the relationship (i.e., correlation⁹) between examinees' scores on their initial attempt and their final attempt.

Table 4
Correlations Between Initial and Final FYLSX Scores
For FYLSX 1st-Time Test Takers

| | | Final Attempt | | |
|-----------------|-----------------|-----------------|------------|------------|
| | | Multiple Choice | Essay | Total |
| Initial Attempt | Multiple Choice | <u>.74</u> | .45 | .66 |
| | Essay | .37 | <u>.46</u> | .46 |
| | Total | .66 | .55 | <u>.69</u> |

While the .46 correlation between an examinee's initial and final scores on the essay portion of the FYLSX can be considered weak (more on this later), the opposite is true for both the scores from MC portion of the exam ($r=.74$) and the combined MC and Essay ($r=.69$). These correlations indicated that there was a very strong tendency that examinees who score poorly on their initial attempt also do so on their subsequent attempt, while the opposite was true for higher scoring students. The data in this table also suggest that the FYLSX Total Scale scores were much more highly related to a student's initial MC Scale

⁹ The correlation is an index ranging from -1.00 to +1.00. It provides a quantitative measure of the relationship between two variables. Correlations less than .5 indicate a "weak" to "moderate" relationship, while those between .5 to .7 are considered "strong" and those above .7 are "very strong".

score ($r=.66$) than to their Essay Scale score ($r=.46$). This finding was also observed in the Klein & Bolus (2010) study reported 12 years ago.

The degree to which initial FYLSX performance was related to examinees subsequent performance can be further illustrated by examining the passage rates, final scores and improvements on a final attempt *stratified by decile ranges* (i.e., each score ranges containing 10% of the sample) of the initial Total Scale scores. These results are presented in Table 5. (Recall that the sample of 4,813 examinees are those that took the FYLSX, failed the exam, and opted to make at least one additional attempt.)

Table 5

**Relationship Between Initial and Final FYLSX Performance:
FYLSX 1st-Time Test Takers who Fail on Their 1st Attempt and
Choose to Retake the Examination**

| Total Initial Score Decile | % Pass | Improvement Over 1st Attempt | | | | | | | | |
|---------------------------------------|---------------|-------------------------------------|--------------|--------------|----------------------------|--------------|--------------|-----------------------------------|--------------|--------------|
| | | Final Average Score | | | Median Score Points | | | % Showing Some Improvement | | |
| | | MC | Essay | Total | MC | Essay | Total | MC | Essay | Total |
| 1st: <=362 | 3% | 181 | 191 | 372 | 19 | 21 | 45 | 72% | 70% | 78% |
| 2nd: 363-399 | 9% | 210 | 217 | 427 | 21 | 20 | 37 | 71% | 67% | 75% |
| 3rd: 400-425 | 12% | 228 | 228 | 456 | 22 | 18 | 38 | 73% | 66% | 74% |
| 4th: 426-446 | 20% | 239 | 238 | 478 | 19 | 19 | 39 | 72% | 66% | 76% |
| 5th: 447-464 | 32% | 252 | 247 | 500 | 25 | 19 | 40 | 78% | 64% | 72% |
| 6th: 465-481 | 39% | 257 | 255 | 511 | 19 | 23 | 45 | 73% | 64% | 71% |
| 7th: 482-496 | 51% | 265 | 265 | 530 | 19 | 24 | 53 | 76% | 68% | 75% |
| 8th: 497-512 | 54% | 269 | 268 | 536 | 15 | 20 | 43 | 71% | 65% | 71% |
| 9th: 513-527 | 66% | 275 | 274 | 549 | 14 | 22 | 38 | 70% | 68% | 73% |
| 10th: >=527 | 76% | 286 | 279 | 565 | 11 | 16 | 27 | 68% | 66% | 76% |
| All | 36% | 246 | 246 | 492 | 18 | 20 | 39 | 73% | 66% | 74% |

Each row in Table 5 represents a score range containing 10% of those who failed on their first attempt (i.e., approximately 480 examinees). For each of those groups, the eventual FYLSX passing rate, final scores by section, and overall score are also presented. To provide insight into how much improvement was generally made across attempts, two statistics were calculated: the median number of “improvement” points (i.e., the level achieved by 50% of the students, approximately 240 examinees) and the percentage of the 480 examinees who experienced at least 1 scale score point improvement from their initial attempt. For example, of the roughly 480 applicants with an initial Total Scale score of between 363 and 399 (i.e.,

the 2nd decile), only 9% passed on their final FYLSX attempt. 50% of those applicants did increase their final Total Scale score by 37 points, and fully 75% realized some improvement in their scores.

From Table 5, we can readily see the relationship between performance on the initial and subsequent FYLSX attempts. Because these calculations were based on students who failed, the top of the score ranges was less than 560 points. The passage rates from Table 5 clearly shows that the lower the initial score, the lower the chances of passing were. Those in the bottom two decile ranges (i.e., below 400 on the initial attempt) had less than a one in 10 chance of passing, *independent of the number of subsequent tries*. To have better than a 50/50 chance of passing, an applicant would have had to initially scored somewhere in the 8th decile (497-512, i.e., 80% higher than all other students).

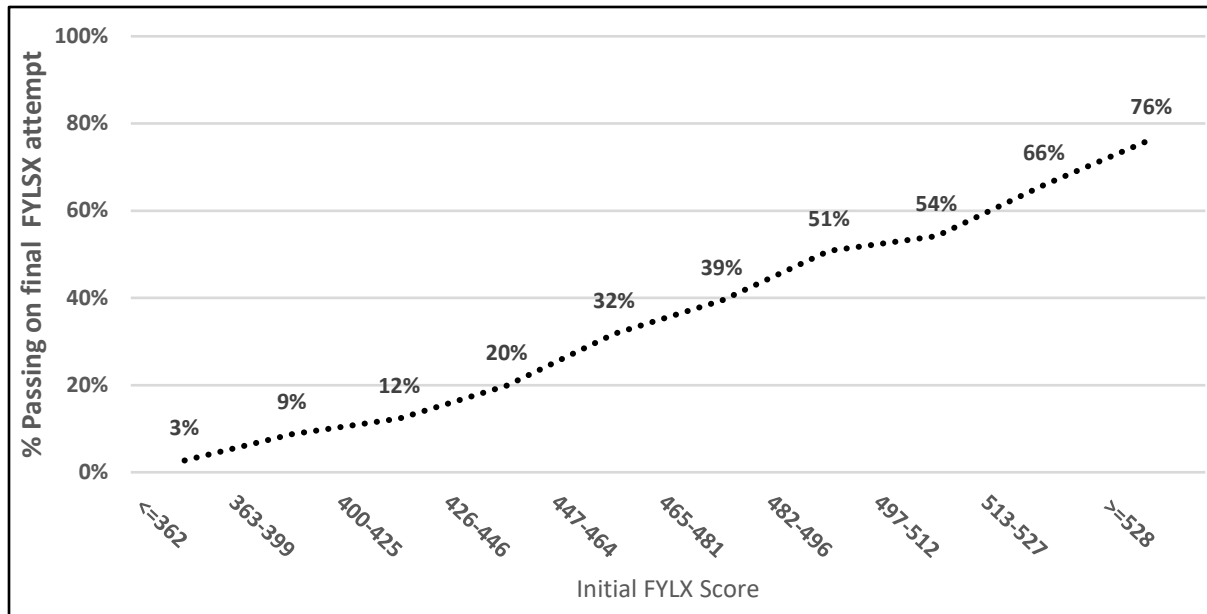
We can also see from Table 5 that scores did improve on subsequent efforts. The median improvement was 39 points between the initial and final efforts. However, the points gained upon further testing did not vary much between the decile ranges. This was also true for the percentage of applicants who experienced any improvement at all (74% overall). Unfortunately, because of these results, examinees with initially low FYLSX scores *simply did not experience enough of a gain in in their scores to improve their chances of passing*. It is unclear from this data whether those students scoring in the lower deciles were qualitatively different (e.g., had less test preparation resources available, were required to work and hence had less study time, etc.) from those who initially scored closer to the passing standard.

One thing we do know is that the differences between the score ranges are not due to perseverance. A separate analysis was conducted to determine the number of re-attempts by students within each of the initial score ranges. We discovered that that the average number of subsequent FYLSX exams attempted by examinees in each of the score ranges was the same for each of the decile ranges (i.e., mean=2.0 attempts). Further, each decile had at least one examinee who has tested 11 times, the maximum number of subsequent attempts possible. Therefore, in the absence of knowing what their eventual chances of passing were, examinees, on average, would tend to make the same number of tries at passing the FYLSX once they have failed, independent of how well they performed on their first attempt.

Figure 5 presents a graphical representation of the relationship between initial performance on the FYLSX and eventual pass rate for those who failed on their first attempt. What becomes readily apparent from the illustration is how the student's chances of eventually passing the FYLSX increased in lockstep with the students' initial performance. For comparative purposes, Table 6 presents the initial and final FYLSX passing rates side-by-side for each of our demographic and legal education groups.

Figure 5

**Relationship between Initial FYLSX scores and
Eventual FYLSX Passing Rate for Examinees Making a Subsequent Attempt**



4. Impact of Legal Education and Demographics on Eventual FYLSX Performance. Beyond initial performance, we examined the degree to which other attributes may have influenced examinees' chances of eventually passing the FYLSX. We did this by developing a series of logistic regression models, expanding those in the Klein & Bolus (2010) report, to include additional demographics¹⁰. The initial model in the series looked only at the initial FYLSX score, and reflected the preceding set of results. The subsequent models then included the demographic and education variables to evaluate their incremental impact¹¹.

As was expected from the results in the preceding section, the first logistical regression which included just initial FYLSX scores was quite strong, resulting in a "c" value of .79. The subsequent

¹⁰ Logistic regression is similar to the linear models discussed earlier, but are appropriate when the outcome takes on dichotomous values (e.g., pass/fail). These models provide the opportunity to simultaneously analyze multiple inputs (e.g., racial/ethnic characteristic, gender) and determine their incremental impact. The strength of the model is determined by the "c" statistic (i.e., "concordance") which measures the degree of agreement between predicted outcome (e.g., pass) and actual outcome. Models with c statistics of .5 or less are considered no better than chance, anything above .7 is considered a good model and .8 or better a strong model.

¹¹ A separate analysis revealed that over 98% of the failing students attended the same school (or program) on their final attempt as they did on their initial attempt.

models, which included all of the student attributes information (i.e., race/ethnicity, gender, and legal education) increased the “c” value of the subsequent models to only .81 (and increase of .02 points).

The details from the logistic regression models presented some very useful insights into the relative impact of student attributes on their likelihood of passing.

To illustrate the impact of student attributes on their eventual FYLSX outcomes, we used parameters from the models to calculate the *estimated eventual passing rates* at three different initial FYLSX score levels, representing the midpoints of the ranges in the top three decile groups. For each specific demographic (e.g., race/ethnicity), we held the other demographic factors constant¹². The results are presented in Table 6.

Table 6
Predicted Eventual FYLSX Passing Rates for Examinees
Failing on Their Initial FYLSX Attempt

| Group | Initial FYLX | | |
|---------------------------------|---------------------|-------------------|-------------------|
| <u>Race/Ethnicity*</u> | <u>504</u> | <u>520</u> | <u>540</u> |
| Asian | 52% | 60% | 70% |
| Hispanic | 43% | 52% | 63% |
| Black | 35% | 43% | 54% |
| White | 43% | 52% | 62% |
| <u>School**</u> | | | |
| ABA | 37% | 46% | 57% |
| Accredited | 36% | 44% | 55% |
| Unaccredited | | | |
| <i>Correspondence</i> | 49% | 58% | 69% |
| <i>Distance Learning</i> | 44% | 52% | 63% |
| <i>Fixed-Facility</i> | 43% | 52% | 63% |

We can see from the results in Table 6, that while initial FYLSX scores are highly predictive of eventual FYLSX scores, after holding them constant, we continued to see differences in passing rates in different racial/ethnic groups (as large as 17%) and between type of legal education (primarily between the Unaccredited schools as a group and other school types). If we consider 1st time FYLSX scores as representing basic ability levels, then these findings would suggest there are other factors operating within the respective demographic groups which impacted the student’s final FYLSX outcome.

¹² * For race/ethnicity, we assumed a male examinee attending a Fixed-Facility institution and for ** School we assumed a Hispanic male. Student gender showed no statistically significant effects.

For any of a number of reasons, not all of the 7,504 1st Time FYLSX takers who failed on their initial attempt opted to re-take the exam. Table 8 presents the eventual pass/fail rates further stratified by their number of attempts, along with their FYLSX scores on their final attempt. The table provides comparative results from the Klein & Bolus (2010) report covering the exam years 1999-2006 (“Previous”) to this study which covers the exam period 2000 to 2016.

Table 8

**Comparison of Final FYLSX Outcomes
Obtained in the 2010 (Previous) Report to Current Report**

| Group | % of Sample | | Multiple Choice | | Essay | | Total | |
|-----------------------|--------------------|----------------|------------------------|----------------|-----------------|----------------|-----------------|----------------|
| | Previous | Current | Previous | Current | Previous | Current | Previous | Current |
| Passed | 44% | 44% | | | | | | |
| 1 Attempt | 30% | 27% | 296 | 296 | 302 | 303 | 599 | 599 |
| > 1 Attempt | 14% | 17% | 284 | 284 | 293 | 292 | 578 | 577 |
| Failed | 56% | 55% | | | | | | |
| 1 Attempt | 27% | 30% | 228 | 223 | 225 | 218 | 452 | 441 |
| > 1 Attempt | 29% | 25% | 221 | 214 | 220 | 212 | 441 | 425 |
| Total | 100% | 100% | 254 | 251 | 256 | 252 | 510 | 503 |

Inspection of the findings are quite remarkable in terms of the results from the two studies. Overall eventual passing and failing rates are virtually identical while the average FYLSX scores in the respective groups differ by no more than a few points. However, we did observe a couple of substantive differences between the outcomes of the two studies. First, while the overall passing and failing rates are the same, in the current study there was a slightly higher pass rate on subsequent attempts along with a corresponding slightly lower fail rate (which we attribute to a longer follow-up period in the current study). We also found slightly lower average test scores among failing applicants, most likely a function of the slight downward trend in FYLSX scores over the recent years. *Interestingly, about 3 out of every 10 applicants who initially sat for the FYLSX (or, more than 1 out of 2 that initially failed) did not attempt again, even though their average scores were really no different (Mean=441) than those who failed but made further attempts (mean=425).* This finding was entirely consistent with that of Klein & Bolus (2010).

The absence of demographic data in the previous report precluded investigating this finding further at the time. Using the available data in the current database, we were able to determine that there were no significant differences in either race or gender between those initially failing applicants who reattempted

vs. those that did not. We did discover that students from Unaccredited law schools were much more likely to make a subsequent attempt upon failing (56%) than those from Accredited (42%) or ABA (36%) institutions. We suspect that this may be due in part to the fact that students in the latter two groups were taking the FYLSX for different reasons than those in the Unaccredited school.

B. What are the GBX outcomes for FYLSX takers?

For most students receiving their legal education within California, their ultimate objective is to eventually pass the GBX and become a practicing lawyer. As we've seen in the data above, only 44% of the 10,340 first-time examinees included in the total sample ultimately passed the FYLSX, and after completion of their education were eligible to sit for the GBX. This section of the analyses focuses on profiling those applicants and discusses how they eventually fared on the GBX.

1. A Profile of Eventual GBX Takers. 3,616 students (35%) from the original cohort of 10,340 subsequently sat for the GBX. These included 3,264 (90.3%) who had passed the FYLSX and a small cohort of 352 (9.7%) who failed the FYLSX but found alternative ways to qualify for the GBX (e.g., transferring to an ABA or Accredited school, special petition, etc.)¹³. The composition of this group of GBX takers is presented in Table 9 on the following along with a comparison to the full group of FYLSX first-time takers.

Table 9 reveals that the sample of students that eventually took the GBX are not dissimilar to the initial population. Looking first at the exam period, we note that the proportion of test takers differ by only a few percentage points, suggesting that the pipeline from the FYLSX to the GBX has remained fairly consistent over the years. This conclusion is further supported by the fact that the passage rates among first-time takers had not varied significantly over that period. We also see a disproportionately larger number of Whites moving on to take the GBX (62% in the GBX sample vs. 54% in the population). This was primarily the result of disproportionately fewer Blacks sitting for the GBX (11% vs 6%). This was not unexpected given the lower likelihood of Blacks eventually passing the FYLSX. And, with respect to legal education, the percentage of students coming from each type of institution when they first sat for the FYLSX is virtually identical to the percentages who will go on to take the GBX. That is, no one type of institution is relatively more successful than another at assisting students to pass the FYLSX and encouraging them to eventually attempt the GBX.

¹³ A separate analysis showed 94% of the students from Unaccredited schools who passed the FYLSX also initially registered to take the GBX as a student from an unaccredited school. This was not the case for the ABA students, where only 40% registered for the GBX as coming from an ABA school. This is not unexpected since the reason that most of these students took the FYLSX was because they were disqualified. The remaining 60% of those students registered as either coming from an Accredited, Unaccredited, or other type of educational program.

Table 9

**Relative Composition of FYLSX Examinees:
Overall Cohort vs. Those Taking the GBX**

| <u>Category</u> | All FYLX 1st Time Takers (N=10,340) | Eventual GBX Takers (N=3,616) |
|------------------------------------|--|--|
| <u>Overall</u> | 100% | 35% |
| <u>Exam Period</u> | | |
| 2000-2003 | 21% | 22% |
| 2004-2007 | 31% | 33% |
| 2008-2011 | 26% | 25% |
| 2012-2016 | 22% | 20% |
| <u>Race/Ethnicity</u> | | |
| Asian | 10% | 11% |
| Hispanic | 11% | 10% |
| Black | 11% | 6% |
| White | 54% | 62% |
| Other | 7% | 6% |
| Missing | 8% | 5% |
| <u>Gender</u> | | |
| Male | 56% | 59% |
| Female | 40% | 38% |
| Missing | 4% | 3% |
| <u>School</u> | | |
| ABA | 6% | 7% |
| Accredited | 6% | 6% |
| Unaccredited | 85% | 83% |
| <i>Correspondence</i> | 23% | 21% |
| <i>Distance Learning</i> | 45% | 44% |
| <i>Fixed-Facility</i> | 18% | 18% |
| Other | 2% | 3% |
| <u>No. of FYLX Attempts</u> | | |
| 1 Attempt | 53% | 62% |
| 2 Attempts | 24% | 20% |
| 3 Attempts | 12% | 10% |
| 4 Attempts | 5% | 4% |
| 5 or more | 6% | 4% |
| <u>Type of Examinee</u> | | |
| Regular | 90% | 89% |
| Disqualified | 9% | 9% |
| Special | 1% | 2% |

2. Initial GBX Performance. Table 10 presents the initial GBX performance of the 3,616 FYLSX takers that sat for the exam. In addition, the table presents the passage rates on the exam under two different standards. The first, “1,440”, represents the passage rates achieved during that period using a 1,440 standard, while the second simulates what the passage rate would have been using the GBX’s current passing standard of 1,390¹⁴. We will refer to this latter standard as the “adjusted passage rate”.

Table 10
Performance of the FYLSX Cohort on Their
Initial GBX Attempt

| FYLX Exam Period | Scale Score | | | % Passing | |
|---------------------|-------------|---------|-------|-----------|-------|
| | MBE | Written | Total | 1,440 | 1,390 |
| 2000-2003 | 1,406 | 1,349 | 1,369 | 29% | 38% |
| 2004-2007 | 1,400 | 1,349 | 1,367 | 30% | 37% |
| 2008-2011 | 1,392 | 1,350 | 1,365 | 29% | 38% |
| 2012-2016 | 1,379 | 1,346 | 1,360 | 28% | 36% |
| All | 1,395 | 1,349 | 1,366 | 29% | 37% |

Table 10 reveals a slight downward trend in the MBE performance (1,406 to 1,379) of FYLSX examinees over the study periods, but no decline for the Essay section. As a result, there was only a nominal change (9 points) in the overall average Total Scale scores. These scores resulted in fairly constant unadjusted GBX passings rates (29% and 37%, respectively) across the different cohorts

When compared to the overall average performance of *all* 148,222 first-time GBX takers between February 2000 and July 2021, we observed that that performance of the FYLSX takers was significantly lower. On average, MBE performance was 80 points lower (1,474 vs 1,395), Written performance was 123 points lower (1,472 vs. 1,349) and Total Scores were 109 points lower (1,475 vs. 1,366). Passage rates were more than twice as low (62% vs. 29%), and while the difference in adjusted passage rates were not quite as large, they were still striking (71% vs. 37%). These differences were actually muted somewhat considering that the performance of the 148,222 GBX takers included the 3,616 FYLSX takers. Later, we will return to the interesting difference between performance on the MBE and Essay GBX.

¹⁴ Since August 2020, the passing standard on the GBX has been 1,390. Before that point, it was 1,440. Because the follow up for FYLSX takers extended to the recent July 2021 exam, we created the “1,440” passing standard to control for the difference. The “1,390” standard was artificially applied to the examinees’ scores, to simulate what might have occurred had that standard been historically applied. It does not account for the potential impact of regrading scores within lower ranges, so the “1,390” passage rates in this report may be slightly understated.

3. Eventual GBX Performance. Similar to their FYLSX efforts, examinees may or may not choose to retake the GBX after they initially fail. As they get closer to their goal of becoming a lawyer, and there are no policy constraints on the number of attempts they can make, it is interesting to discover what their final outcomes were.

As Table 10 indicates, 1,049 FYLSX takers (29%) passed the GBX on their initial attempt. While there were 2,567 eligible to retake the GBX, fully 2,130 (83%) chose to do so. This meant that 437 (17%) examinees who failed on their initial attempt made no additional attempts at the GBX. This is almost half the 30% rate of those who never made a subsequent attempt at passing the FYLSX upon their initial failure.¹⁵ We may conclude that as students get closer to their ultimate goal of becoming a lawyer, they are more likely to persevere.

Table 11 provides data on the final GBX scores for the 2,130 examinees who did choose to repeat the GBX after failing on their first attempt.

Table 11

**Final GBX Performance for those repeating the GBX
Along with Information on Their Improvement**

| FYLX Exam Period | Final Average Scale Scores | | | Improvement Over 1st Attempt | | | | | |
|---------------------|----------------------------|---------|-------|------------------------------|---------|-------|----------------------------|---------|-------|
| | | | | Average Score Points | | | % Showing Some Improvement | | |
| | MBE | Written | Total | MBE | Written | Total | MBE | Written | Total |
| 2000-2003 | 1,413 | 1,355 | 1,376 | 80 | 73 | 76 | 77% | 71% | 80% |
| 2004-2007 | 1,404 | 1,357 | 1,375 | 73 | 68 | 71 | 76% | 71% | 78% |
| 2008-2011 | 1,402 | 1,368 | 1,382 | 74 | 76 | 77 | 76% | 74% | 79% |
| 2012-2016 | 1,410 | 1,372 | 1,391 | 86 | 86 | 89 | 81% | 75% | 86% |
| All | 1,407 | 1,362 | 1,380 | 77 | 74 | 77 | 77% | 72% | 80% |

Examinees who repeat the GBX clearly improved by an average of 77 points over the 16-year study period. Within the 2012-2016 cohort, virtually all applicants (86%) saw some level of improvement in their scores. Correlational analysis also demonstrated that the examinees who performed relatively well on their initial

¹⁵ While the window for tracking these examinees through their GBX efforts was quite large, it is possible that some FYLSX takers from the latter years (e.g., 2015 and 2016), might yet choose to attempt a GBX. Analysis showed that 95% of the FYLSX takers had made their initial GBX attempt within 5 years of passing the FYLSX. In terms of our cohort, that would be the July 2021 GBX administration for those passing in October 2016. Thus, while there may be a few within our study cohort who have not attempted a GBX, the chances are quite small.

attempt also tended to perform relatively well on their subsequent attempt, especially on the MBE ($r=.67$). The relationship for the Written section were slightly weaker ($r=.50$), due most likely to the lower reliability of that section of the GBX. The correlation for GBX Total scale scores was .63.

Table 12 presents the overall picture of GBX passage rates (both initial and final) after all subsequent attempts for the sample of 3,616 FYLSX takers who sat for the GBX.

Table 12
Initial and Final GBX Passage Rates
Under Original and Alternative Passing Standards

| FYLSX Exam Period | Initial | | Final | | Difference | |
|------------------------------------|----------------|-------------|--------------|-------------|-------------------|-------------|
| | 1440 | 1390 | 1440 | 1390 | 1440 | 1390 |
| 2000-2003 | 29% | 38% | 57% | 62% | 28% | 24% |
| 2004-2007 | 30% | 37% | 56% | 62% | 26% | 25% |
| 2008-2011 | 29% | 38% | 59% | 64% | 30% | 26% |
| 2012-2016 | 28% | 36% | 59% | 62% | 31% | 26% |
| All | 29% | 37% | 57% | 62% | 28% | 25% |

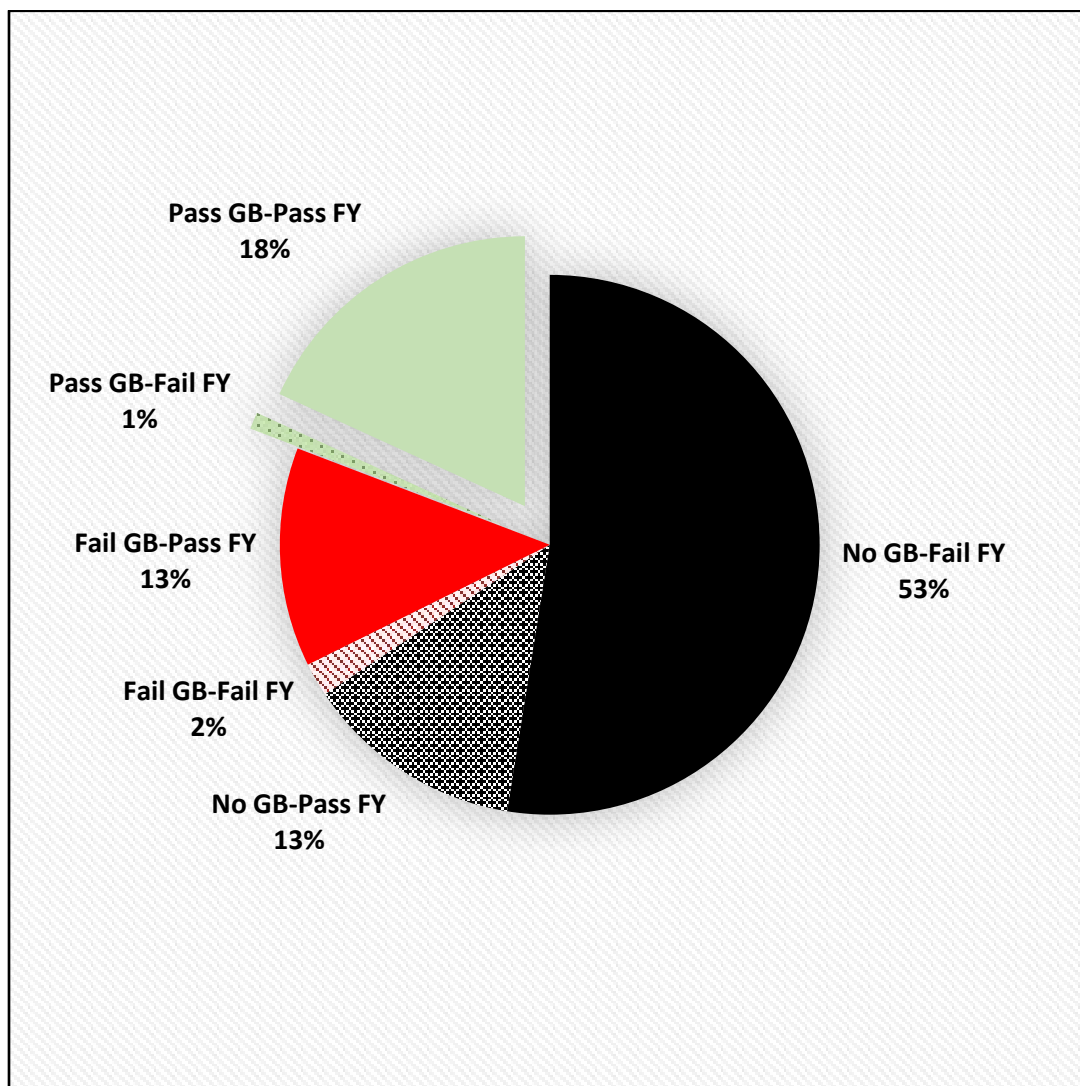
Across all years, we observed that the final passage rate almost doubled from the initial to the final attempt (29% to 57%). Under the alternative passing standard of 1390, the passage rate rose from 37% to 62%. These results were quite similar to those found in the Klein & Bolus (2012) report, which showed an eventual GBX passage rate of 50% and 59% for cohorts of students from Unaccredited and Accredited law schools that were tracked for 6 subsequent administrations. We would expect the rates in this report to be slightly higher as the follow-up were longer than 6 GBX administrations, and the current sample includes some examinees who eventually graduated from ABA schools.

A graphical summary of the eventual outcomes of our starting FYLSX cohort of 10,340 first-time takers is captured in Figure 6 on the following page. Based on a GBX passing standard of 1,440 we see that less than 20% of the initial sample of 13,430 eventually passed the GBX. 18% of the sample passed through a “traditional” pathway of first passing the FYLSX, while 1% of the sample never did pass the FYLSX but found an alternative path to taking and passing the GBX. As noted previously, slightly more than 65% of the total sample never attempted to take the GBX, including the 13% who were eligible because they had passed the FYLSX. Application of the adjusted 1,390 passing standard only slightly impacted the Passing GBX

portion of the pie by increasing the rate for passing both exams section by 2% and likewise lowering the “Fail GBX-Pass-FY” by the corresponding 2%.

Figure 6

**Eventual GBX Outcomes for
The Initial Cohort of FYLSX First-Time Takers**



4. The Relationship of FYLSX to GBX Performance. Clarifying the relationship between performance on the FYLSX to the GBX performance could provide students with valuable information that could aid in a decision to continue their legal education. This is one of the stated objectives of the FYLSX. Towards that end, we conducted a series of analyses designed to provide that information.

Table 13 presents the correlations of initial and final FYLSX scores, with the GBX Total Scale scores for the sample of 3,616 that sat for both exams.

Table 13
Correlations between FYLSX and GBX Scores:
Initial and Final Attempts

| FYLSX Section | Initial FY Scores | | Final FY Scores | |
|----------------------|--------------------------|------------------|------------------------|------------------|
| | Initial GBX | Final GBX | Initial GBX | Final GBX |
| MC | .59 | .50 | .51 | .44 |
| Essay | .43 | .37 | .31 | .28 |
| Total | .58 | .50 | .49 | .44 |

The results in Table 13 reveal a relatively strong relationship between FYLSX and GBX performance, but the relationship is not uniform across sections. Scores on the MC section (either 1st or final attempt) correlated much more strongly with GBX scores than did scores on the Essay section. Actually, the FYLSX MC scores correlated as well with GBX scores as did the Total FYLSX scores. This finding, which replicates the results in Klein & Bolus (2010), suggests the Essay scores on the FYLSX not only added little to the predictive value of the FYLSX, but may actually reduce it.

In fact, the “true” relationship between the FYLSX and GBX is more likely higher than what is reported in Table 13. First, we note that the GBX measures a broader skill set and more subject matter areas than the three subject matter areas of the FYLSX. Additionally, we note that the correlations in the table are “attenuated” (i.e., reduced) since only FYLSX takers with scores greater than 540 are allowed to take the test; the actual range of the FYLSX extends far below that score. Once this range restriction is statistically accounted for, the true correlation between the FYLSX & GBX total scores increased from .58 to .70. We conclude from these observations that the scores from the FYLSX, particularly the MC and the Total Scale are an effective predictor of eventual GBX performance.

We next examined the relationship between FYLSX scores and GBX *passage rates*. Similar to the approach discussed earlier, we used logistic regression to predict the probability of passing the GBX (at first attempt and final attempt) using initial and final FYLSX Total Scale scores. We used the Concordance statistic “c”, described earlier, to evaluate the efficacy of the model. Models were constructed for both the 1,440 and

1,390 passing standards. We also compared the actual and predicted passing rates from the models as another means of assessing model efficacy.

The models, whether they were based on initial or final FYLSX or GBX performance, were all quite effective, with “c” statistics ranging from .70 to .79, attesting again to the predictive validity of the FYLSX. To further examine the efficacy of the models, we looked further into the model with the lowest “c” value (“c” =.70), i.e., the model using student’s final FYLSX Total Scale Score to predict final GBX outcome after all attempts. We reasoned those other models with higher “c” values would yield results that were as good or better.

Using the regression parameters from the logistic model, the probability of passing was calculated for each student based on their final FYLSX score. We then categorized the FYLSX Total Scale scores into intervals of 20 points each, beginning with 540 (the lowest score that a student could have gone into reappraisal). We then calculated the average actual and predicted passage rate for all the students within each of the ranges. The results are provided in Table 14.

Table 14
Prediction of Students Final GBX Outcomes
Based upon Final FYLSX Score Ranges

| | | | 1440 Standard | | 1390 Standard | |
|--------------------|----------|-----------------------------------|-------------------------|----------------------------|-------------------------|----------------------------|
| <u>FYLSX Score</u> | <u>N</u> | <u>% of all FYLSX test takers</u> | <u>Actual Pass Rate</u> | <u>Predicted Pass Rate</u> | <u>Actual Pass Rate</u> | <u>Predicted Pass Rate</u> |
| 541-560 | 725 | 7.0% | 41% | 39% | 47% | 45% |
| 561-580 | 820 | 7.9% | 49% | 50% | 55% | 56% |
| 581-600 | 589 | 5.7% | 60% | 61% | 64% | 67% |
| 601-620 | 434 | 4.2% | 70% | 71% | 77% | 76% |
| 621-640 | 275 | 2.7% | 79% | 79% | 84% | 83% |
| 631-660 | 198 | 1.9% | 88% | 85% | 89% | 89% |
| 661-680 | 102 | 1.0% | 89% | 90% | 91% | 92% |
| 681-700 | 60 | 0.6% | 92% | 93% | 95% | 95% |
| > 700 | 45 | 0.4% | 98% | 98% | 98% | 98% |

Within all ranges, the FYLSX scores were remarkably accurate in predicting actual GBX final passage rates. This extended from the bottom intervals, where over 1,500 students scored, to the upper intervals where only about 100 students from the original sample scored. In only a few of the ranges did the difference

between the actual and predicted rates differ by more than 2%. Also, we can see that model worked equally well for both passing standards.

A closer inspection of the score intervals revealed insights into how the likelihood of passing the GBX changes. We can see that students who scores fell *between 541 and 580 on their FYLSX exam were predicted, on average, to have less than a 50/50 chance of eventually passing the GBX*. After that point, the chances of passing are predicted to improve, on average, by about 10% for every 20-point increase in FYLSX scores. Above 680, students' chances of passing were more than 9 out of 10. However, there were only about 100 students (1%) who scored this high¹⁶.

5. The influence of Student Characteristics and Law School Education on GBX Performance.

The models discussed above were developed without consideration of the demographic characteristics or law school education of students. To determine the degree to which those characteristics might mediate the probability of passing, we conducted further analyses. In each model, we incrementally added a single attribute to the original model (containing only the final FYLSX Total score). Adding gender as the first student attribute in the new models, did nothing to improve prediction; indicating that males and females performed equally well. The addition of race/ethnicity as the next attribute slightly improved the model's "c" value from .70 to .71, while the final inclusion of the law school type (both at the point of taking the FYLSX and last attempt at the GBX – noting that these institutions could be different) led to an overall "c" value of .74. These improvements in the predictive value of the models suggested that both race/ethnicity and educational institution will have an incremental impact on GBX performance, above and beyond how students perform on the FYLSX.

We can evaluate the size of the effect in two ways: first in terms of "relative odds" and then in terms of differences in predicted passage rates. Regular odds simply state the ratio of success to no success. For example, if a successful outcome occurs 3 out of every 4 times, then the odds of success would be 3-to-1. The "relative odds" is a way of evaluating the difference in odds of two or more groups. If one group's

¹⁶ Note: Based on the correlational pattern that we observed in Table 13, we chose to develop an additional set of logistic regression models. In these models, FYLSX MC Scale scores were used instead of the FYLSX Total scores. The resulting "c" statistics from those alternative models were only .01 points lower, while the differences between the actual and predicted GBX passage were only slightly higher (i.e., a maximum difference of 5% in the lower score ranges). This finding underscored the relative importance of the FYLSX's MC section.

odds are 6-to-1, and another group's odds are 2-to-1, then the relative odds were be 3-to-1, meaning that one group is 3 more times likely of achieving success than the other.

After calculating the odds of passing the GBX for each gender, racial/ethnic group and educational institution, we calculated the relative odds within each of those groups. As mentioned above, we found no impact of gender on GBX total score, and this resulted in a relative odds ratio of 1-to-1 for males vs. females. With respect to race, comparing the performance of Whites to the other racial/ethnic groups we obtained relative odds ratios of 1.3 for Whites vs. Asians, 1.6 for Whites vs. Hispanics and 1.9 for Whites vs. Blacks. Holding FYLSX performance constant (i.e., students with the same score) then, Whites are 1.3 times as likely as Asians to pass the GBX, 1.6 times as likely as Hispanics and 1.9 as likely as Blacks.

Since law school attended could change over time, we looked at the effects of law school type at two time points, when they first sat for the FYLSX and then when they registered to take the GBX. We found that students attending ABA-approved schools when they first took the FYLSX were about twice as likely to eventually pass the GBX as those from Unaccredited Correspondence or Distance Learning programs (Relative odds ratios=1.9 for each), 1.3 times more likely than those attending Accredited institutions, and interestingly about $\frac{1}{2}$ as likely as those from Unaccredited Fixed Facilities (We note here that most ABA students at that point were taking the FYLSX for disciplinary reasons).

The differences in the chances of passing between ABA schools and other institutions were more pronounced when the graduating institution (i.e., their institution of record on the GBX application) was considered. A student from an ABA school was 3.3 times more likely to pass the GBX than one from an Accredited school, 9.9 times as likely as one from a Fixed Facility Unaccredited school, and 5.4 and 4.5 times as likely as ones graduating from a Distance Learning or Correspondence Unaccredited schools, respectively. We can infer from these findings, that while the initial FYLSX performance has a small impact on the chances of passing the GBX, subsequent educational experiences have a much more profound mediating influence.

An alternative method of illustrating the effects of legal education and student demographics on the likelihood of GBX passing is to apply the parameters from the logistic regression model to actually estimate the expected probabilities. In Table 15 we demonstrate a few alternative scenarios in which we compare the expected GBX passage rates at three, fixed score levels of the FYLSX (at the bottom of the reappraisal range, at the passing standard, and well above the passing standard). In deriving these estimations, we held constant other factors. For example, in the race models shown in Table 15, we show results for a

male student from an Unaccredited Fixed Facility law school; in the school models, we show results for a White female graduating and registering for the GBX from the same institution. Absolute percentages in Table 15 would be higher or lower if other demographic combinations were used, but the relative differences between the racial/ethnic groups or school types would remain the same.

Table 15

**Prediction of Students' Final GBX Passage Rates (1440)
For Different Racial/Ethnic Groups & Law Schools at Alternative FYLSX Scores*¹⁷**

| <u>Group</u> | <u>Final FYLSX Score</u> | | |
|--------------------------|--------------------------|------------|------------|
| <u>Race/Ethnicity*</u> | <u>540</u> | <u>560</u> | <u>580</u> |
| Asian | 36% | 47% | 58% |
| Hispanic | 32% | 42% | 53% |
| Black | 28% | 38% | 49% |
| White | 43% | 54% | 65% |
| <u>School**</u> | | | |
| ABA | 83% | 88% | 92% |
| Accredited | 37% | 48% | 59% |
| Unaccredited | | | |
| <i>Correspondence</i> | 25% | 35% | 45% |
| <i>Distance Learning</i> | 31% | 41% | 52% |
| <i>Fixed-Facility</i> | 36% | 47% | 58% |

Looking first at the Race/Ethnicity portion of the table, we see that White students had a predicted 7% to 8% higher chance of passing the GBX than Asians, 11% to 12% higher chance than Hispanics and a

¹⁷ For Race/Ethnicity predictions, we assumed a Male graduating and registering for the GBX from a Fixed Facility Law school. For the School predictions, we assumed a White Female graduating and registering for the GBX from the same school.

14% to 16% higher chance than Blacks, depending upon the particular FYLSX score level. These findings should not be interpreted as simply a function of the examination's inherent bias against any of the groups. Rather, other factors related to exam performance may be confounded with race/ethnicity (e.g., exam preparation, work requirements, study resources, language barriers). Similarly, we see that attending and graduating from an ABA school has a strong relationship to GBX success. Over 80% of ABA students with an initial FYLSX score of 540 would be predicted to pass the GBX, climbing to over 90% if their initial score was 580. The predicted rates from the Accredited and Unaccredited schools are less than ½ that of the ABA schools, though the rates rise much more quickly as their FYLSX scores increase. These findings are markedly similar to those found by Klein & Bolus (2010).

C. Modifying the Structure and Scoring of the FYLSX.

The preceding findings substantiate both the reliability and predictive validity of the FYLSX as well as the exam's capacity to serve as a guidepost for determining whether students from Unaccredited (and Accredited) law schools will eventually have sufficient skills to pass the GBX and become licensed lawyers. Furthermore, the study analyses have identified areas for potential improvements to the examination. These improvements could lead to reduced testing burden on examinees as well as efficiencies for the State Bar's overall FYLSX testing program. We believe that three elements of the FYLSX including the essay section, the reappraisal process, and the passing standard, all present opportunities for change.

1. The Essay Section.

The 4-question essay section makes up one-half of the FYLSX, both in terms of the amount of testing time as well as its contribution to the overall score. In having an Essay section, developers are required to create new questions for each administration. Some have argued that writing is an essential legal skill and should be assessed in any examination process related to a 1L curriculum. Yet results presented earlier in Table 13 revealed that scores from the essay portion were only weakly related to GBX scores, with a correlation statistic as low as $r=.28$ with GBX essay scores. From a psychometric perspective, a major source of the weak relationship emanates from the low reliability of the essay portion. An inspection of reliability data from the historic FYLSX technical reports clearly illustrates this.

Appendix 2 presents the section and total score reliability of the FYLSX indices for 30 administrations dating back to June 2007. The Appendix illustrates that reliability of the MC section scores have been consistently quite high, averaging .89. These levels come close to meeting the upper reliability standards of a high-stakes examination (even in the absence of an essay portion). In fact, they are comparable to a

100-item version of the current NCBE MBE examination. In comparison, the reliability of the 4-Question Essay section of the FYLSX has varied significantly over the years, ranging as low as .50 to as high as the .70's on recent administrations. The average across the 30 administrations was .63, fully .26 points lower than the MC section average. Clearly, the reliability of the FYLSX and the subsequent predictive value for GBX performance could be improved if some adjustments were made.

At least two options are available for affecting improvement:

1) Adjusting the Weighting of the Multiple Choice and Essay Section. While there is little documentation on the rationale for equally weighting the two sections of the FYLSX during scoring, the approach most likely evolved from the notion that is commonly held regarding the GBX; i.e., the written section is a more valid assessment of legal knowledge, and as such should be weighted more heavily or, at a minimum, equal to a multiple-choice test. When the current version of the FYLSX was first developed, the written portion of the GBX actually was given 65% of the scoring weight while the MC portions (i.e., the MBE) was given only 35%.

When the configuration of the GBX was modified in 2017 to accommodate two days of testing, psychometric analyses indicated that the weighting needed to be adjusted i.e., weighting of the MBE needed to be increased to 1/2 in order to maintain the same historical levels of overall GBX score reliability. As can be observed in Appendix 2, by using the current 50/50 weighting schema, the overall FYLSX reliability has averaged a respectable .85, dropping as low as .80 in October 2014 (when the Written section reliability was .50) to as high as .91 on the most recent October 2021 administration. On *all but the most recent administration, however, the overall reliability of the FYLSX was actually below that of the MC section.*

If including an Essay portion on the FYLSX is considered essential, one possible solution could be adjusting its relative weighting. To evaluate the impact of changing scoring weights of the respective sections, we examined two alternative scoring schemas. In each schema, the MC section was given more weight than the Essay section; in the first, the weighting was 2/3 to 1/3 and in the second, the weighting was 3/4 to 1/4¹⁸. As can be seen in the latter columns of Appendix 2, the overall FYLSX score reliability levels went up, on average by .04 and .05 points, respectively. Under the 3/4 MC weighting scheme, the average reliability was .90, in the upper levels of reliability standards for a high-stakes examination. ***Thus, an***

¹⁸ The overall reliability of the test was estimated using the methods in Wang & Stanley (1970).

argument can be made to modify the section weighting based on the potential for improved score reliability.

Improvements in reliability lends partial evidence in support of a FYLSX scoring modification. However, if such a change has an adverse impact on the overall FYLSX passage rates or a differential impact on selected demographic groups, then it might not be justified. Earlier sections of this report showed that under the current system, there were minimal to no differential impacts on FYLSX performance attributable to group membership after other factors were taken into consideration. Therefore, any changes to the scoring would have to meet this standard as well.

To evaluate this, two additional FYLSX scores were recalculated for the 10,340 first-time FYLSX takers in our sample and subsequently compared to the current scoring. The two additional simulated scores were based on the different MC weightings suggested above. One score was calculated as $2 \times (.67 \times \text{MC Scale score} + .33 \times \text{Essay Scale score})$ while the other was calculated as $2 \times (.75 \times \text{MC Scale score} + .25 \times \text{Essay Scale score})$. Each of the students' simulated scores were then evaluated to determine if they fell above or below the current 560 passing standard. We then made a series of comparisons between the actual (before reappraisal) and simulated outcomes.

Analysis of the results revealed, as expected, a high degree of relationship between the simulated and actual Final FYLSX Total Scale score ($r=.99$ and $r=.97$ for the $2/3$ and $3/4$ weighting, respectively). Average mean simulated Total Scale scores differed from actual Total Scale scores by no more than .7 points (on the 0 to 800 scale) while their score spreads (i.e., Standard Deviations) differed by no more than .5 points. The overall passing rate was identical between the actual and simulated $3/4$ MC weighting (34.1%) and differed by only .3% between the actual and simulated $2/3$ MC weighting (34.1% vs. 34.4%). A small percentage of examinees actually changed pass/fail status under the simulated conditions (9.8% under the $3/4$ MC weighting; 6.6% under the $2/3$ MC weighting); roughly equal percentages changed status, switching from pass to fail and from fail to pass.

While small shifts in the final examinees' outcomes might be expected (especially for those scoring around the 560 standard), the concern would be whether those shifts were more favorable for one group or another. A series of X^2 tests conducted for gender, racial/ethnic group and law school type indicated, however, that none of the specific groups benefited more from the re-weighting schema than any other group.

2) Removing the Essay from the FYLSX. The more extreme response to the reliability issues of the Essay portion of the FYLSX would be to completely eliminate this section of the exam. This option would eliminate the need for the biannual question development process as well as the need to grade examinees' answers. The reappraisal phase would be eliminated as well, since pass/fail decisions would be based solely on multiple-choice scores which involve no subjectivity in scoring. Additionally, testing time would be limited to a ½ day with the possibility of implementing more than twice-a-year testing.

In their prior discussion of this option, Klein & Bolus (2010) posited that:

“The strongest statistical arguments for deleting the essay section are (a) this section is unnecessary because FYLSX essay scores do not contribute to the prediction of GBX scores or pass/fail status provided by the multiple-choice section alone, and (b) FYLSX essay scores are not reliable enough to be useful for diagnostic purposes...differences between individual questions are too unstable to be trusted (P.10)”

As a counterpoint, however, the authors suggested that retaining the essay might encourage candidates to improve their writing skills.

To further investigate the impact of eliminating the essay completely in this study, we conducted two additional logistical regression analyses. In these analyses we substituted the student's final FYLSX MC Scale score for their Total Scale score as the Independent Variable in the prediction of the likelihood of passing the GBX (at both the 1440 and 1390 passing standard). The resulting models yielded “c” statistics ($c=.70$), identical to the ones reported earlier based on the Total Scale score. We conclude from this finding the Essay portion of the FYLSX continues to contribute little in the screening for eventual success on the GBX.¹⁹ ***Therefore, from a psychometric standpoint, we see removing the Essay section of the FYLSX completely as a viable option.***

2. Reconsideration of the Reappraisal Phase.

The philosophy supporting inclusion of a reappraisal phase is the recognition that some degree of error exists in all measurement, and that it is costly to students to make a false negative decision; i.e., failing a student when that student is worthy of passing.²⁰ There is little documentation in the State Bar's archives, however, that provides the rationale for either the Reappraisal Phase itself, or how it was implemented

¹⁹ We might add here that, as an option, if the State Bar were to substitute an additional afternoon of 100 MC items (making for a 200-item test...similar in length to the MBE), we would expect the reliability of the MC portion of the test to climb to $r_{tt}=.95$, with a further increase in its predictive power.

²⁰ We've calculated the Standard Error of Measurement (SEM) for the FYLSX at about 34 points, which would indicate that the reappraisal range is about ½ a SEM.

(i.e., grader's deciding on an outright FYLSX pass/fail decision as opposed to adjusting scores on the Essay portion and recomputing scores). Nor is there any information indicating how the reappraisal score range (i.e., 540 to 560) was chosen.

Inclusion of a Reappraisal Phase does come at a cost. It requires additional grading resources to be used, and generally the more highly trained ones. Additionally, as a consequence of where the current reappraisal range lies within the overall score distribution for the FYLSX, more students score within that range than any other 20-point score range on the exam, which creates a heavy workload for these graders. And reappraisals can only be conducted at the end of general grading and scoring, thus extending the length of time to complete the process and delaying the release of results to all students, not just those with scores within the reappraisal range. Given these considerations, we reasoned that it would be worthwhile to look more closely at the cost/benefit of the Reappraisal Phase.

Across the 34 administrations in the study time frame, 1,161 (11.2%) of the 10,340 first time FYLSX takers had their essay questions reassessed in Reappraisal. The percentage of students going into the Reappraisal phase has varied by only a few percent over that time.

We examined the likelihood of passing the FYLSX once a student made it into Reappraisal. Table 16 presents the passage rates (in two-point intervals) of students scoring within the Reappraisal range. Inspection of Table 16 indicates that overall, an examinee who made it into the Reappraisal grading phase had more than a 90% chance of passing the exam, while those with scores over 550 had a greater than 95% likelihood. The likelihood of passing decreased somewhat toward the bottom of the range, however, 73% of those in the 541-542 range still passed.

Further analysis revealed that only 62% of students undergoing reappraisal eventually took the GBX. Overall, of those that did sit for the GBX, 40% eventually passed although the passage rates varied between 30% and 50% depending on the score interval. Passage rates were not directly correlated with FYLSX scores, however. That is, as the FYLSX scores increased from 540 to 560, the pass rates did not increase correspondingly.

These findings suggest that a final reappraisal phase may not be justified. We see a few options for improvement. First, the passing standard could simply be lowered by 20 points to 540, while simultaneously eliminating the Reappraisal phase. Over the 16-year study period, this solution would have had the net effect of passing 91 more students (i.e., 1161-1070 or .8% of the total sample of 10,341), or an average of between 5 and 6 more students per year. Had this change been implemented, it would

have had minimal impact on the GBX as only 91 more students would have been eligible to sit for the GBX, and based upon the rates of who eventually will take the GBX, we estimate that only about 60 would have opted to do so.

Table 16

**Eventual FYLSX Passage Rates for Examinees
Going in FYLSX Reappraisal**

| <u>FYLSX Score</u> | <u>Total Examinees</u> | <u>Passing Examinees</u> | <u>% Pass at that Score</u> |
|------------------------|----------------------------|------------------------------|-------------------------------------|
| < 560 | 109 | 109 | 100% |
| 558 | 129 | 127 | 98% |
| 556 | 154 | 153 | 99% |
| 554 | 140 | 138 | 99% |
| 552 | 138 | 133 | 96% |
| 550 | 122 | 107 | 88% |
| 548 | 95 | 83 | 87% |
| 546 | 88 | 72 | 82% |
| 544 | 87 | 72 | 83% |
| 542 | 79 | 58 | 73% |
| 540 | 20 | 18 | 90% |
| Total | 1161 | 1070 | 92% |

An alternative option that would have a more profound impact would be to maintain the current passing standard of 560 while eliminating the reappraisal process entirely. Over the study period, this option would have had the net effect of failing an additional 1,161 candidates, lowering the eventual FYLSX passage rate from 44% to 33%, (assuming no student reattempted). If the same reattempt rate for failing students was applied along with the subsequent pass rates identified in Table 8, we would expect an eventual FYLSX passage rate to increase to 35%; a passage rate that is still 9% lower. We believe that both alternatives warrant consideration by the State Bar.

3. Reconsideration of a FYLSX Passing Standard.

The current passing standard for the FYLSX has been set at 560 for over 30 years, and was originally set to ensure that a roughly equivalent proportion of examinees would pass (relative to historical rates) after scoring changes were made to the examination in the late 1990's. During the modifications of the FYLSX

that led to its current version, two standard setting studies were conducted. The first involved external panelists comprise of deans and faculty members from ABA, California Accredited and Unaccredited law school, and the second used experienced FYLSX graders; Klein (1998). The outcomes indicated regular attorney graders of the FYLSX would have set the standard at 574 while the external panelists evaluation would have set the passing score at 538. Additionally, panelists from the Unaccredited schools were statistically more likely to pass answers that panelists from the other two school types were not²¹. The Committee of Bar Examiners opted not to change the standard from 560, and the passing standard has not been revisited since that time.

We believe that the statistics presented in this report offer some insights into the adequacy of the current standard given that the stated objective of the FYLSX is to assist in screening students who could eventually become qualified to take, and subsequently pass the GBX. To consider the appropriateness of the 560- standard, we utilized the logistic regression models applied in Table 14. Recall that these models used students' final FYLSX scale scores to predict eventual success on the GBX. The overall model was considered quite strong as evidenced by its "c" value of .70. Data in Table 15 illustrated the consistency between actual bar passage rates (at both a 1,390 and 1,440 standard) and the rates predicted in the models. The score range in Table 14 extended from the bottom of the Reappraisal range to scores greater than 700.

While the models were initially developed based on actual scores of applicants within the 540-to-800 point range, it is possible with some caution, to apply the resulting equation to examinees scoring below that range. In other words, some insight might be gained into expected GBX passage rates if the decision were made to lower the cut score on the FYLSX.²² Table 17 presents the estimations that resulted in applying the models to final FYLSX score ranges below 540.

For each FYLSX score interval in the leftmost column, we determined the number of FYLSX examinees from our original sample of 10,340 that fell into that interval, along with the cumulative count as we moved lower in the score distribution. We then applied the logistic regression model to estimate the likelihood of a student passing within each range. We calculated the average of those probabilities to arrive at an expected passage rate, along with number of passers (multiplying the rate by number of

²¹ Note that except for a very lenient evaluation of one question, Klein indicated that panelists from the Unaccredited schools would have set the standard at 560.

²² We should note that it is impossible to be 100% confident that the nature of the relationship that we established between passage rates and FYLSX scores would remain the same for score ranges below 540, we would assume that for scores close to 540 would not deviate too severely.

examinees who fell into that score range). We maintained a running count of estimated additional passers, and then estimated the overall GBX passing rates for the entire group that eventually took the GBX²³. We made these calculations based on both a passing standard on the GBX of 1,440 and 1,390. Note that score ranges closest to the current passing standard are in 10-point increments while those further away (lower than 500) are in 20-point increments.

Table 17
Projected GBX Passage Rates
For Examinees Scoring Below 540
On Their Final FYLSX Attempt

| | | | 1440 | | | | 1390 | | | |
|--------------------|-------------------------------------|--|---------------------------------|-------------------------------------|--------------------------------------|---------------------------------------|---------------------------------|-------------------------------------|--------------------------------------|---------------------------------------|
| Score Range | Examinees Within Score Range | Cumulative Examinees Within Score Range | Pred Pass Rate W/N Range | Additional Examinees Passing | Cumulative Additional Passing | Expected Overall GBX Pass Rate | Pred Pass Rate W/N Range | Additional Examinees Passing | Cumulative Additional Passing | Expected Overall GBX Pass Rate |
| 531-540 | 212 | 212 | 32% | 68 | 68 | 56% | 37% | 78 | 78 | 61% |
| 521-530 | 229 | 441 | 27% | 62 | 130 | 54% | 32% | 73 | 151 | 59% |
| 511-520 | 282 | 723 | 23% | 65 | 195 | 52% | 27% | 76 | 227 | 57% |
| 501-510 | 296 | 1019 | 19% | 56 | 251 | 50% | 23% | 68 | 296 | 55% |
| 481-500 | 638 | 1657 | 15% | 96 | 347 | 46% | 17% | 108 | 404 | 50% |
| 461-480 | 664 | 2321 | 10% | 66 | 413 | 42% | 12% | 80 | 484 | 46% |
| 431-460 | 628 | 2949 | 7% | 44 | 457 | 39% | 8% | 50 | 534 | 42% |
| 421-440 | 590 | 3539 | 4% | 24 | 481 | 36% | 5% | 30 | 563 | 39% |
| 401-420 | 511 | 4050 | 3% | 15 | 496 | 34% | 3% | 15 | 579 | 37% |

As an example, we can look more closely at the projected effect of lowering the passage score to > 530 (531-540)²⁴. Out of our sample of 10,340 first-time FYLSX takers, there were 212 examinees within that range across the 17-year period. Based on our logistic regression model, we estimated that 32%, or 68 additional examinees, would have passed the FYLSX using a 540-passing score. Adding 68 to the “passers” count and the additional 212 examinees to those who would now sit for the GBX, we estimated that the overall *group passage rate* would fall slightly from 57% (see Table 12) to 56%, or to 61% if the current GBX standard of 1,390 was applied. If the standard were to be dropped to > 520, then we would estimate that

²³ In deriving these estimates, we assumed all students in this range would opt to take the GBX. As such, the resulting rates may be overestimates since not all FYLSX passers have gone on to take the GBX.

²⁴ We did not assume any Reappraisal, and based on the analyses above that examined Reappraisal, we reason that this would be a good starting range.

130 more examinees would have passed (68+62) and the overall eventual GBX passage rate for these examinees would fall by another 2% to 54%. If the FYLSX passing score was dropped all the way down to 400, an additional 4,050 examinees could be expected, *et ceteris paribus*, to sit for the GBX, while the expected overall eventual GBX passage rate for the group would fall to 34%. We can also see that the expected percentage of examinees who are estimated to pass within each of the FYLSX score intervals is estimated to get progressively smaller the further the interval deviates from the current standard.

In reviewing these numbers, it is important to keep in mind several factors. First, as noted in the footnote and demonstrated in earlier tables, not all examinees who pass the FYLSX opt to take the GBX. Secondly, the counts of additional examinees who would be expected to pass if the FYLSX standard was lowered are based upon 17 years of examination data (34 administrations). If applied to a single administration, only 1 to 2 additional examinees per score interval would have passed, if the FYLSX passing standard was dropped all the way down to 400, (which is 160 points from the current standard!).

Finally, it is essential to understand that these rates and counts are simply the results of a mathematical model. It is unknown what the behavior of examinees scoring within these ranges would be if they were allowed to pass. It is unclear if they would decide to continue their legal education just because they passed the FYLSX. The model does not, and cannot, factor in what grades these students have earned in law school, though we can safely assume that they would be lower than their fellow students who actually passed the FYLSX. Thus, any interpretation of these projected passing rates most likely should be considered as upper bounds. Finally, setting a passing standard must involve considerations beyond actual and projected passage rates, and should rely on the judgment of experts. Whether the eventual GBX passage rate is 60% or 34%, this statistical analysis simply presents a context in which to evaluate whether the FYLSX is functioning as it was designed, and whether the passing standard has been set appropriately.

REFERENCES

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

Hosmer DW, Lemeshow S: Applied Logistic Regression. 1989, John Wiley & Sons, New York, NY.

Klein SP (1982) "An analysis of the relationship between initial score and eventual pass/fail status on the California Bar Examination". A report prepared for the Committee of Bar Examiners of the State Bar of California.

Klein SP (1998) "Panelists and reader judgements regarding the passing score on the FYLSX". A report prepared for the Committee of Bar Examiners of the State Bar of California.

Klein SP (1999) "How passing the FYLSX is related to passing the GBX". A report prepared for the Committee of Bar Examiners of the State Bar of California.

Klein SP & Bolus RE (2010) "How well does the FYLSX Predict Success on the GBX". A report prepared for the Committee of Bar Examiners of the State Bar of California.

Klein SP & Bolus RE (2012) "Eventual passing rates on the California Bar Exam by school type for six cohorts of first-time takers beginning with February 2007 takers." A report prepared for the Committee of Bar Examiners of the State Bar of California.

Wang, M and Stanley, J (1970) 'Differential weighting: A review of methods and empirical studies. Review of Educational Research, 40, pp. 663-705.

Appendix 1

Initial and Eventual FYLSX Passage Rates By Demographic and Legal Education Groups

| | Initial Status | | | | | Final Status | | | | |
|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------|
| Category | 2000- 2003 | 2004- 2007 | 2008- 2011 | 2012- 2016 | All Years | 2000- 2003 | 2004- 2007 | 2008- 2011 | 2012- 2016 | All Years |
| Overall | 30% | 28% | 26% | 25% | 27% | 45% | 47% | 41% | 43% | 44% |
| <u>Race</u> | | | | | | | | | | |
| Asian | 27% | 27% | 27% | 20% | 26% | 46% | 50% | 46% | 42% | 46% |
| Hispanic | 21% | 17% | 19% | 18% | 18% | 36% | 39% | 34% | 35% | 36% |
| Black | 11% | 6% | 11% | 8% | 9% | 23% | 20% | 22% | 24% | 22% |
| White | 39% | 36% | 35% | 34% | 36% | 54% | 56% | 51% | 53% | 54% |
| Other | 21% | 20% | 17% | 23% | 20% | 39% | 40% | 34% | 41% | 38% |
| Missing | 18% | 22% | 10% | 13% | 16% | 23% | 27% | 16% | 20% | 22% |
| <u>Gender</u> | | | | | | | | | | |
| Male | 33% | 30% | 30% | 27% | 30% | 50% | 50% | 45% | 45% | 48% |
| Female | 28% | 27% | 24% | 23% | 25% | 41% | 46% | 39% | 42% | 42% |
| Missing | 15% | 6% | 3% | 11% | 9% | 20% | 7% | 8% | 13% | 13% |
| <u>School</u> | | | | | | | | | | |
| ABA | 16% | 27% | 25% | 13% | 21% | 37% | 43% | 40% | 27% | 38% |
| Accredited | 21% | 12% | 10% | 10% | 14% | 30% | 30% | 21% | 20% | 26% |
| Unaccredited | 33% | 30% | 28% | 26% | 29% | 47% | 49% | 43% | 44% | 46% |
| Correspond. | 38% | 31% | 33% | 29% | 32% | 50% | 48% | 46% | 46% | 47% |
| Distance | 37% | 33% | 27% | 27% | 31% | 52% | 53% | 43% | 46% | 49% |
| Fixed | 20% | 18% | 20% | 20% | 20% | 37% | 39% | 40% | 37% | 38% |
| Other | 19% | 20% | 21% | 25% | 22% | 36% | 41% | 36% | 48% | 41% |
| <u>Type</u> | | | | | | | | | | |
| Regular | 32% | 29% | 27% | 26% | 28% | 47% | 49% | 42% | 44% | 46% |
| Disqualified | 11% | 17% | 16% | 12% | 14% | 26% | 32% | 26% | 22% | 28% |
| Special | 56% | 43% | 50% | 28% | 45% | 68% | 67% | 79% | 55% | 67% |

Appendix 2

Historic and Projected FYLSX Reliability Estimates Under Current and Alternative Section Weightings

| <u>Exam</u> | <u>Multiple Choice</u> | <u>Essay</u> | <u>Total Score</u> | | |
|-------------|----------------------------|--------------|--------------------|-------------------------------|-------------------------------|
| | | | <u>Current</u> | <u>2/3 MC - 1/3 Essay</u> | <u>3/4 MC - 1/4 Essay</u> |
| Jun. 2007 | .89 | .67 | .86 | .89 | .90 |
| Oct. 2007 | .90 | .57 | .83 | .89 | .90 |
| Jun. 2008 | .90 | .58 | .84 | .89 | .90 |
| Oct. 2008 | .90 | .52 | .82 | .88 | .90 |
| Jun. 2009 | .89 | .56 | .83 | .88 | .89 |
| Oct. 2009 | .90 | .55 | .82 | .88 | .90 |
| Jun. 2010 | .90 | .64 | .86 | .90 | .91 |
| Oct. 2010 | .89 | .55 | .83 | .88 | .89 |
| Jun. 2011 | .91 | .63 | .86 | .90 | .91 |
| Oct. 2011 | .88 | .67 | .86 | .89 | .89 |
| Jun. 2012 | .88 | .66 | .85 | .89 | .89 |
| Oct. 2012 | .89 | .51 | .80 | .87 | .89 |
| Jun. 2013 | .89 | .64 | .85 | .89 | .90 |
| Oct. 2013 | .89 | .64 | .85 | .89 | .90 |
| Jun. 2014 | .86 | .67 | .85 | .88 | .88 |
| Oct. 2014 | .88 | .50 | .80 | .86 | .88 |
| Jun. 2015 | .90 | .60 | .84 | .89 | .90 |
| Oct. 2015 | .89 | .56 | .82 | .88 | .89 |
| Jun. 2016 | .88 | .64 | .85 | .89 | .89 |
| Oct. 2016 | .87 | .72 | .87 | .89 | .89 |
| Jun. 2017 | .90 | .67 | .87 | .90 | .91 |
| Oct. 2017 | .90 | .66 | .86 | .90 | .91 |
| Jun. 2018 | .89 | .62 | .84 | .89 | .90 |
| Oct. 2018 | .90 | .63 | .85 | .90 | .91 |
| Jun. 2019 | .90 | .73 | .88 | .91 | .91 |
| Oct. 2019 | .91 | .63 | .86 | .90 | .92 |
| Jun. 2020 | .90 | .65 | .86 | .90 | .91 |
| Oct. 2020 | .90 | .74 | .89 | .91 | .92 |
| Jun. 2021 | .92 | .75 | .90 | .92 | .93 |
| Oct. 2021 | .90 | .74 | .91 | .92 | .92 |
| Average | .89 | .63 | .85 | .89 | .90 |

Appendix 3

Logistic Regression Parameters from Model Predicting Final GBX Passage

| <u>Parameter</u> | <u>Raw</u> <u>Regression</u> <u>Estimate</u> | <u>Stdzd</u> <u>Regression</u> <u>Estimate</u> |
|----------------------------|--|--|
| Intercept | -10.96 | |
| Final FYLX Score | .02 | .47 |
| Gender | | |
| Male | -.06 | -.02 |
| Female | -.08 | -.02 |
| Race | | |
| Asian | -.42 | -.08 |
| Hispanic | -.62 | -.10 |
| Black | -.79 | -.10 |
| White | -.13 | -.03 |
| Other/Missing | -.76 | -.10 |
| Initial Institution | | |
| ABA | -1.05 | -.13 |
| Accredited | -1.33 | -.15 |
| Unaccredited | | |
| Fixed Facility | -.60 | -.14 |
| Distance Learning | -1.71 | -.47 |
| Correspondence | -1.69 | -.35 |
| Final Institution | | |
| ABA | 1.84 | .17 |
| Accredited | .65 | .09 |
| Unaccredited | | |
| Fixed Facility | -.45 | -.11 |
| Distance Learning | .15 | .04 |
| Correspondence | .35 | .07 |

The standardized weights have been added to the table to provide the reader with a direct comparison of the *relative* contribution of each of the variables in the model.