2018

Technical Report on the Texas Bar Examination

A review of the procedures, results and psychometric characteristics of the 2013 through 2017 examinations.

Roger Bolus, Ph.D.
Research Solutions Group
6/29/18

SUMMARY

This report presents the outcomes of a technical review of the statistical and psychometric properties of the Texas Bar Examination (TBE) during the five-year period from 2013 to 2017. This review was commissioned by the Texas Board of Law Examiners (TBLE) to support decision-making regarding the content and administration of the TBE.

Over the course of the study period, the configuration of the TBE remained unchanged. It consisted of four sections: The 200-item Multistate Bar Examination (NCBE), an equated and scaled multiple choice test developed by the National Conference of Bar Examiners (NCBE) used by all but one U.S. State; an Essay section comprised of 12 questions written and scored on a 25-point scale by TBLE members; a short answer Procedure and Evidence (P&E) section prepared by the TBLE, unique to Texas that contains 40 questions on Criminal and Civil Law, each graded on a 5-point scale; and a Multistate Performance Test (MPT) developed by the NCBE, graded on a 0 to 6-point scale by TBLE members. Raw scores on the non-MBE section are converted to a common mean and standard deviation and then scaled to the MBE. The sections scale scores are subsequently weighted (40% MBE, 40% Essay, 10% P&E, 10% MPT) and added to form a single, compensatory Total Scale Score (TSS) which is placed on a 1000-point scale.

Examinees pass outright if they achieve a TSS of 675 or greater, fail if they score 668 or less; or have all written answers re-evaluated a second time if they score between these scores. Points can only be added during the re-grade process. After re-grade, a TSS of 675 or more passes; below that fails.

The statistical and psychometric analysis of the ten administrations studied revealed the following:

- Recalculation of all raw and scale scores for the 21,000+ examinees tested confirmed that the TBLE is accurately following all published calculation specifications.

- Raw and scale scores on each of the test sections regularly demonstrated acceptable levels of variation and have highly consistent and symmetric score distributions; a desirable property contributing to both score reliability and decision consistency.

- Across the July administrations, we observed only the slightest difference in the average Total Scale Scores of examinees repeating the examination for 2nd, 3rd, 4th or 5th attempt (653 to 657). On February administrations, a similar pattern was observed though 2nd time-takers did score +13 scale score points higher than 3rd and 4th time-takers (667 vs 654).

- The correlations among test sections, a key factor in examination reliability, were generally consistent across administrations. In July, the relationships between the MBE and the Essay and P&E averaged .64 and .63, respectively, while the Essay and P&E correlated .67. The MBE/Essay correlations are consistent with those observed in other states with a 12-question test. The weakest and least consistent correlations were observed between the MPT and the other sections (averaging r=.36 and .38 with the MBE and Essay, and .29 with the P&E).

This pattern is not unexpected given that the MPT section receives only a single 0 to 6 score, thereby reducing the reliability of the MPT component.

- Overall, 3.1% and 4.1% of all examinees had their answers reread over the study period on the July and February administrations, respectively. When in re-grade, 99% of examinees earned extra score points on the Essay section as compare to 63% and 44% on the Civil and Criminal P&E sections, respectively. Only 13% on the examinees in re-grade earned extra score points on MPT. As a result of the re-grade process over the 5-year period, the overall bar passage rate increased by 3.8% and 3.0% in February and July, respectively. If an applicant made it into re-grade during that period, they had a 94% chance of passing, and a 100% chance of passing with an initial score of 673 or 674, before re-grading.

- Anchored by the very high reliability of the MBE and strong reliability of the 12-question essay test, the overall reliability of the TBE averaged .89 and .90 for the five February and five July administrations, respectively. These levels exceed the baseline standard of a .85 reliability for a high stakes licensing test.

Overall, the combination of stable relationships over time, and the moderate correlations between sections suggest that the TBE, through use of multiple measures and multiple formats, is consistently measuring a set of diverse legal content areas and skill domains in a consistent fashion. Any decisions to develop alternative configurations of the TBE (e.g., modifying the weighting of the respective sections, dropping sections, or modifying the examination length) would need to be evaluated against the strong results of the current structure.

A. INTRODUCTION

This report presents the outcomes of a technical review of the statistical and psychometric properties of the Texas Bar Examination (TBE) during the five-year period from 2013 to 2017. This review was commissioned by the Texas Board of Law Examiners (TBLE) to support decision-making regarding the content and administration of the TBE.

In response to the issues raised by some stakeholders, Texas is considering a broad array of questions regarding the exam including, among others:

- Should the examination continue to include the Multistate Bar Examination (MBE);
- Should the essay portion be restructured;
- Should the weighting of the different sections of the exam be modified;
- Should the duration of the exam be reduced to two days; and
- Should the Texas Board of Law Examiners (TBLE) transition to use of the National Conference of Bar Examiners (NCBE) Uniform Bar Examination (UBE)

Any modification to the existing structure of the TBE could have far reaching consequences, and requires consideration of psychometric, as well as, administrative factors. Previous studies conducted by Klein & Bolus (2004, 2008) established the psychometric foundations for the current examination. Subsequently, an audit of the 2013 testing cycle verified that scoring procedures were applied per specification and provided other technical data on the exam's administration (Klein & Bolus, 2014). Other than reports of passing rates, however, no other technical information on the TBE has been available since that time.

Many of the questions addressed by the Supreme Court Task Force on the Texas Bar Examination are best considered in context with the type of technical data included in the 2014 Klein & Bolus report. In order to support future decision making, the TBLE requested an update and expansion of the prior study to include a review of a full five-year period covering the ten February and July administrations of the TBE from 2013 to 2017. This report presents the outcome of this review in accordance with the technical documentation requirements laid out in the Standards for Educational and Psychological Testing (2014), the de facto guideline for high stakes license testing.

The report addresses multiple topics. First, because it has been fully documented elsewhere, we briefly describe the current structure, grading and scoring procedures of the TBE. [1] We then examine the distributional properties of the raw and scale scores of examinees. We report on passage rates and examine the outcomes of Texas' re-grade procedures. We present data on the reliability of the exam's separate components and composite scores used for making pass/fail decisions. In presentation of the results, we examine trends and stability of results over the five-

---

[1] An overview of the examination is provided on the TBLE's website: https://ble.texas.gov/bar-exam-general-instructions

year period from 2013 to 2017, and we study the similarities and differences that exist between February and July administrations.

The study findings that follow provide the statistical foundation for an assessment of how well the TBE has functioned during the period of study and serves as a key reference point for the consideration of potential modifications to the examination.

B. THE TEXAS BAR EXAMINATION

1. Examination Structure and Administration

Over the five-year period covered in this study, the structure and administration of the TBE has remained unchanged. The TBE consists of four sections administered over two and a half days.

Multistate Performance Test (MPT). The MPT is a constructed response task developed by NCBE and designed to measure an applicant's ability to complete a life-like task that would be required of a first-year lawyer. This section of the TBE is administered on the morning of the first day of testing and applicants are given 90 minutes to complete it. The test usually consists of a series of tasks involving a standardized set of realistic case materials that the applicant must review and assimilate before constructing their response. Currently, all but 14 states administer one or two MPTs as part of their examination. The test is considered by many to be the most valid portion of the bar examination, because it most closely approximates the work of a lawyer.

Procedure and Evidence Test (P&E). The P&E is a 90-minute test administered after the MPT on the first morning of the TBE. The P&E consists of two separate sections: (a) The Civil Law test covering topics on Texas civil procedure and evidence, and (b) the Criminal Law test coving topics on both Federal and Texas criminal procedure and evidence. Each section consists of 20 short answer (no more than five written lines) questions that have been constructed by members of the TBLE. A new set of questions is created for each administration. No other state bar examination uses this form of testing.

Multistate Bar Examination Test (MBE). The MBE is a 200-item multiple choice test developed by NCBE that is designed to test legal knowledge and analysis skills in seven legal doctrinal areas. The MBE is administered on the second day of testing, with 100 items administered during a 180-minute morning session, and the balance administered in a 180-minute afternoon session, for a total of six hours of testing. Currently, 175 of the items are considered "live" and contribute to applicants' performance scores on the MBE. The remaining 25 items are considered "experimental" with resulting data used to evaluate the quality of the questions and adequacy for use in subsequent test administrations. A portion of the 175 items have appeared on previous versions of the test and are used for equating purposes (see the discussion of test scoring below). For security purposes, multiple forms of the test, each with the same items but different orderings, are used. All states administering a bar examination in the U.S. (except Louisiana) use the MBE.

Essay Test (Essay). On the third and final day of the TBE, a 12-question essay test is administered. Questions are administered in two 6-question blocks; a morning block and afternoon block, each allotted 180 minutes of testing time. In general, essay questions are designed to assess legal knowledge, analysis, reasoning and sometimes, application skills. Each question is expected to require an average of 30 minutes to complete, but applicants can budget their time in any manner that they choose within each block. The test questions cover any of nine areas of law (including cross-over topics). The current content domains for the questions were implemented in July 1999. The questions are constructed, edited and revised by members of the TBLE and professional staff. Unique questions are developed for each test administration and essay questions are not reused. Questions can be answered either in hand-written form or typed on laptop computers.

2. Examination Grading and Scoring

Similar to all other state bar examinations, each section of the TBE is graded/scored separately and the resulting section scores are mathematically combined to form an overall score.

MBE. Scoring on the MBE is performed by NCBE. For each examinee, a "raw" score is calculated based upon the number of correct responses. That score is then "equated" or "scaled" and placed on a 200-point scale. The process of equating involves comparing the scores of current test-takers on "anchor" items given on previous test to the performance of applicants from previous tests. NCBE then uses that information to adjust the scores on the remaining items, thereby controlling for differences in the difficulty levels of newer items on the test. Thus, a resulting equated score on the current examination reflects the same level of ability as that same equated score on any previous administration of the MBE. Through this process, any differences in scores across administrations can be attributed to differences in the proficiency level of examinees, as opposed to differences in how hard or easy the respective test items were. The process yields a *Scaled MBE Score* for each examinee.

P&E. Each of the 20 items on the Civil and Criminal sections of the P&E are scored on a 0 to 5 scale using an analytical scoring guide. Higher scores represent a more complete and accurate answers. The scores on the 20 items are then summed to form a "raw" score on each test (maximum score=100). The raw scores on each test are then "converted" to a scale that has a mean of 100 and standard deviation (i.e., standard score spread) of 10. This conversion process is undertaken so that when the two scores are combined to form a total P&E score, mathematically, each section is of equal difficulty and carries equivalent weight.[2] Then to control for any differences in difficulty of the P&E over time, the converted P&E score is linearly *scaled* to the same distribution as the MBE (which, itself represents scores that have been equated over time), yielding a *Scaled P&E Score*.

---

[2] When combining scores from two different tests, the effective "difficulty" and "weight" that each test has is based upon the mean and standard deviation of the test's score distribution. The wider the distribution, the greater the weight. By standardizing the mean of each of the P&E's test section scores (i.e., 100), it controls for any differences in difficulty between the two tests; and by applying the same standard deviation (i.e., 10) it insures each portion carries the same weight.

MPT. The MPT is scored on a single 0 to 6-point scale (in 1-point increments) by one of two trained and previously calibrated graders using the analytic scoring guide provided by NCBE. Similar to the P&E, the Raw MPT score is then scaled to the MBE distribution to yield a *Scaled MPT Score*.

Essay. Each of the 12 essay questions are graded by a separate grader using a pre-established scoring guide developed for the question. Each essay is graded on a 1 to 25-points raw score scale.   In order to control for differences in essay difficulty, grader leniency and score spread, the raw scores on each question are then converted to a distribution with a common mean of 100 and standard deviation of 10. These scores are then summed to form a Converted Total Essay. For the reasons mentioned above, the Converted Total Essay is then also linearly scaled to the MBE Scale Score distribution to form the *Scaled Total Essay Score*. Linear scaling to the MBE for all sections was first adopted for the February 2008 examination.

Combined Total Scale Score. With scores for all four sections of the exam on the same scale of measurement (i.e., the 200-point MBE Scale), a weighted Total Scale Score is calculated. The TBLE weights the MBE and Essay sections equally at 40% each, and gives10% weighting to the MPT and P&E sections. To get the scores on a 1000-point scale, the mathematical calculation of the Total Scale Score reflecting these weights are:

Total Scale Score = (2 x MBE Scale) + (2 x Essay Scale) + (.5 x MPT Scale) + (.5 x P&E Scale)

The use of this method of total score calculation is considered "Compensatory" in nature as it allows for better performance on one section to compensate for poorer performance on other sections. The Total Scale Score is converted to a whole number based on traditional rounding rules.

3. Pass/Fail Decisions and Re-grading

Phase 1 Decisions.   Based upon the value of the Total Scale Score, an applicant can pass, fail, or have his/her written answers re-evaluated in a re-grade process. This determination is based on the following criteria for the Total Scale Score:

1. >= 675   -> Pass
2. <= 668   -> Fail
3.   669 to 674   -> Regrade

The Re-grade Process and Phase 2 Decisions. If an examinee enters the re-grade process, all their constructed responses (P&E, MPT and Essay) are reread by a second grader. That grader sees the original score and has the option of adding raw score points but cannot subtract points. As a result, an examinee's scores can only increase in re-grade, but cannot decrease.

After re-grade is completed, the examinees raw scores are recalculated. Using the same conversion and scaling equations used during the initial scoring, the applicant's converted and scaled section scores are also recalculated, as is their weighted Total Scale Score. Based upon that score a final pass or fail decision is made according to the following criteria:

1. >=675   -> Pass
2. < 675   -> Fail

The passing standard of 675 is equivalent to a 135 on the MBE scale of measurement. The 135 value is the pass/fail criterion used by 15 other states[3] and is also the most common. There is significant variation, however, across the country in (a) the use of a re-grade process, (b) the methods for how an examinee gets into re-grade, and (c) the actual grading/scoring of re-graded answers.

---

[3] Among the remaining states, roughly one-half use a higher standard, and half use a lower standard.

C. EXAMINATION RESULTS

Currently the TBLE's statistical reporting consist of the tabulation of overall bar passage rates with a breakdown of those rates by various examinee characteristics (e.g., repeater status, law school, etc.). In this section, we extend that reporting by including salient statistical and psychometric results considered germane to the documentation of the quality of the TBE. The Executive Director requested that we study five years of actual examination results. This is a sufficiently long enough period to evaluate recent trends in the results and the impact that those trends may have had on the quality of the examination itself.

Study Data. Data for the study was provided by the technical staff of the TBLE. Ten data files were prepared, one for each administration February and July administration for 2013 to 2017. The files contained one record per applicant and included raw scores (the original grading/scoring scale) on each of the examination sections[4] both before and after re-grade, as well as converted scores and scale scores (see below for an explanation). For nine of the ten examinations, scores on each of the short answer questions on the Procedure and Evidence portion were included in the files[5]. Finally, the number of times the applicant had previously attempted to pass the exam was provided as well. No demographic data on examinees or the law school they attended was included, however.

The data files were created in EXCEL format and transferred to a secured site accessible only to TBLE technical and Research Solutions Group (RSG) staff. The data files were translated into SAS format (a statistical analysis database product), verified and combined into a single analytic database that was used for all analyses.

Information on the structure of the examination, grading procedures and the scoring of the examination discussed in the previous section was obtained from the TBLE website, other available documentation and conversations with the Executive Director. Finally, comparative data on national statistics were obtained from publications provided by NCBE as well as from the author's work with other state jurisdictions.

Table 1 contains the counts of test taker records for each of the examinations. Over the five-year period there were slightly more than 21,200 examinees, averaging about 1,200 test takers on the February administrations, and close to 3,000 examinees in July. As with other U.S. jurisdictions, first-time test takers made up the preponderance of examinees on the July administrations, averaging 84%. However, there has been a steady decline since 2013 when first-timers made up 90%. That proportion had dropped to 80% in 2017. There has been a corresponding and steady increase in the proportion of applicants repeating the examination on both the February and July examinations, rising 12% and 10% respectively between 2013 and 2017. As discussed later in the report, this shift has been a function of the change of bar passage rates over that period.

---

[4] Raw scores are not available for the MBE
[5] Because of database conversions, the TBLE did not have detailed, P&E question-specific scores on the July 2016 P&E examinations.

Data Quality and Accuracy Scoring Procedures. To validate the quality and accuracy of the records in the data file, we performed two sets of analyses.

We first compared the count of records in the transferred data to the annual February and July exam statistics published on the TBLE website, Both the aggregate examinee counts as well as the number of first time takers and repeaters were identical for all ten tests under study, insuring that we were studying the entire cohort.

Table 1

Number of First Time and Repeating Applicants
Sitting for the Texas Bar Examination
2013 through 2017

	Total	**First Time Taker**		**Repeater**	
	N	**N**	**%**	**N**	**%**
February					
2013	1,185	732	61.8%	453	38.2%
2014	1,152	742	64.4%	410	35.6%
2015	1,333	725	54.4%	608	45.6%
2016	1,433	689	48.1%	744	51.9%
2017	1,253	610	48.7%	643	51.3%
Average	1,271	701	55.6%	602	47.2%
July					
2013	3,023	2,709	89.6%	314	10.4%
2014	2,929	2,548	87.0%	381	13.0%
2015	2,987	2,512	84.1%	475	15.9%
2016	2,975	2,393	80.4%	582	19.6%
2017	2,959	2,349	79.4%	610	20.6%
Average	2,975	2,503	84.1%	472	15.9%

We then examined the individual data elements available in the data sets. Since we had access to both raw and scale scores, we were able to independently replicate all the calculations (i.e., conversions, scaling, weighting and score combinations) that were performed for the 2014 audit report previously referenced. As for that study, *every calculation was verified for every test taker over the five-year period.* We concluded that Texas has continued to perform its scoring per specification and we had a valid dataset to perform analyses.

The remainder of this section reports on the results of our analyses.

1. Descriptive Statistics.

An objective of licensing examinations is to adequately spread out applicants on a continuum of performance and ensure that the full range of the measurement scales are utilized. Toward that end, our first analyses of the TBE examined the distributional properties of each of the test sections. Distributional properties can be summarized using a variety of statistics. Given the number of examinations being studied and the number of tests, we opted to focus on five common statistics:

1. Mean (*Mn* or *Avg*)   - A common measure of central tendency representing the arithmetic average of all scores

2. Standard Deviation (*Sd.*) – The standard measure of score spread representing the averages of the differences of all scores from the overall average

3. Median (*Md*) – Another measure of central tendency representing the score on the distribution for which 50% of the scores fall above and 50% below.

4. 25th Percentile (*25%ile*) - The score on the distribution of all scores at which 25% of test takers score <u>below</u> and 75% perform <u>above</u>.

5. 75th Percentile (*75%ile*) - The score on the distribution of all scores at which 75% of test takers score <u>below</u> and 25% perform <u>above</u>.

Knowledge of the score distribution statistics is useful for many reasons. First, it provides a single summary measure on how the typical, or average, individual performed on a given administration. Second, it provides a reference point for the location on the distribution where any given score is situated relative to all other scores. Third, it provides information relative to the overall shape of the distribution. All things held equal, more symmetric, well spread out distributions are preferred on a license test. When section scores are combined (such as on the TBE), it is preferable to have these types of distributional properties on the respective sections, as they lead to more reliable overall scores. Finally, distributional statistics are useful for looking at year-over-year shifts, and for investigation of potential causes of these changes.

Tables 2 and 3 present the distributional statistics for initial raw scores on the essay section and the Criminal and Civil sections of the P&E exams, while Tables 4 and 5 present those same statistics for the MBE and Total Scale Scores. The statistics were calculated separately for each of the ten administrations, as well as for all 6,356 February test-takers and all 14,873 July test-takers[6] combined.

---

[6] We did calculate similar statistics for the MPT scores. We observed little to no differences in the 25th percentile (3 points), 75th percentile (4 points) or Sd (1.1 points) across any of the administrations. The overall average MPT score was only .1 points higher in July than in February. For the sake of space, we did not table these results.

Looking first at the raw score distributions we observed that for both the Essay section (Table 2) and each of the P&E tests (Table 3), both measures of central tendency (i.e., the Mean and Median) are virtually identical, differing by no more than a few points on the 300-point Essay raw point scale and less than 1 point on the 100-point Criminal and Civil Law P&E tests. Additional, the 25th%ile and 75th%ile points are equidistant from central points of both the Essay and P&E tests. These results suggest a highly symmetric score distribution.

While the respective Means (and Medians) do vary year-over-year on both test sections, we notice that the Standard Deviations (Sd) remain markedly similar. On the Essay section the Sd's differ by a maximum of only three points in both February and July; on the Criminal portion of the P&E they differ by two and three points over the five years on the February and July administrations respectively, and by three and two points on the February and July administrations of the Civil portion. The average Standard Deviation on each of the ten P&E tests were each 14 raw score points, except for the February Civil section that averaged 15 points.

Given these findings, the resulting underlying raw score distributions can be characterized as symmetrical and well-conditioned for combining with other score distributions from other test sections. Further, the differences in raw mean scores across the five years suggest potential differences in either test difficulty and/or applicant performance. Thus, the scoring methodology employed by the TBLE to scale these raw scores to an equated scale score distribution, such as the MBE, is quite appropriate and necessary.

Table 2

Distributional Statistics for the
Initial Raw Essay Scores
On the Texas Bar Examination
2013 through 2017

	25th %ile	Md	Ave.	75th %ile	Sd
February					
2013	158	174	173	190	26
2014	175	195	192	212	28
2015	160	176	174	190	25
2016	166	184	181	199	26
2017	136	153	152	171	27
All Years	156	176	174	194	29
July					
2013	178	193	191	207	24
2014	161	178	176	193	24
2015	163	181	178	196	24
2016	153	169	168	185	24
2017	142	162	160	180	27
All Years	158	177	175	194	26

Table 3

Distributional Statistics for the
Initial Raw Criminal and Civil P&E Scores
On the Texas Bar Examination
2013 through 2017

	Criminal					Civil				
	25th %ile	**Md**	**Ave.**	**75th %ile**	**Sd**	**25th %ile**	**Md**	**Ave.**	**75th %ile**	**Sd**
February										
2013	55	66	65	75	14	58	67	65	75	14
2014	59	68	67	77	14	53	64	62	74	16
2015	55	63	62	71	12	50	59	57	67	13
2016	53	63	62	71	14	68	75	73	80	11
2017	49	56	56	64	12	50	58	58	67	13
All Years	53	63	62	72	14	55	65	63	74	15
July										
2013	57	65	64	73	11	61	71	69	79	13
2014	60	68	67	75	11	55	64	62	71	13
2015	44	53	52	61	13	53	63	61	70	14
2016	55	65	64	74	14	51	58	58	67	12
2017	59	57	56	76	12	55	63	63	73	13
All Years	52	61	61	70	14	55	64	63	72	14

Table 4 presents the distributional data for the MBE portion of the TBE. Based on national statistics provided by the NCBE, it is well-established that the distributions of MBE scale scores are highly symmetric and normally configured. An inspection of Table 4 suggests that this is the case for the Texas-specific scores as well.

The Mean and Median scores on every exam are no more than one point apart on any single administration and identical across all years. Further, we see that 25th percentile and 75th percentile points are equidistant from central points of both the February and July Exams. The Standard Deviations, which match closely with national results, are identical for all February administrations and all July administrations, with one exception. The slightly lower Standard Deviation on the February administration, relative to July, reflects the higher concentration of applicants repeating the examination in February.

As was the case in states across the county, a downward trend in TBE examinees' performance on the MBE over the five-year period is apparent: There was a seven point drop in February mean scores between 2013 and 2017 (½ Sd) and a four point drop in July (slightly more than ¼ Sd). Both decreases are slightly smaller than what was observed nationally, however (ten points and five points in July and February).

Finally, Table 6 presents the distributional statistics for the TBE Total Scale Scores. The weighting and combining of well-conditioned distributions from the various test sections result

in an equally symmetrical distribution of overall scores. The Means and Medians of each administration vary by no more than five points on the 1000-point Total Scale Score distribution. Both quartile points are equidistant from the Means (across all exams, roughly 38 points in February and 46 points in July). Additionally, the respective February and July administrations' Standard Deviations vary across the years by no more than 5 points.

Table 4

Distributional Statistics for
MBE Scale Scores
On the Texas Bar Examination
2013 through 2017

(*With inclusion of U.S. Average for comparative purposes)

|	25th %ile	Median	Texas Avg.	U.S. Avg.*	75th %ile	Sd
February						
2013	131	139	140	143	149	14
2014	130	139	139	139	148	14
2015	128	136	136	137	145	14
2016	125	135	135	134	144	14
2017	124	133	133	134	143	14
All Years	127	136	136	137	146	14
July						
2013	134	145	145	147	155	15
2014	130	141	141	143	151	15
2015	128	138	139	140	149	15
2016	130	141	141	140	152	16
2017	131	141	141	142	152	15
All Years	131	141	141	142	152	15

Table 5

Distributional Statistics for the
Final Total Scale Scores
On the Texas Bar Examination
2013 through 2017

	25th %ile	Md	Ave.	75th %ile	Sd	% Passing
February						
2013	667	701	701	739	55	74%
2014	658	695	693	736	58	68%
2015	649	684	682	717	57	60%
2016	640	680	675	712	58	56%
2017	630	667	666	704	60	48%
All Years	647	685	683	721	59	
July						
2013	687	729	724	769	63	82%
2014	661	705	703	747	62	71%
2015	653	696	693	738	63	66%
2016	660	707	703	751	67	71%
2017	662	708	705	752	65	72%
All Years	663	709	706	752	65	72%

The decreases over the years observed in the MBE Score Averages are reflected directly in the decreases in both the Total Scale Score Averages and the percentage of applicants passing the TBLE. Figure 1 illustrates the relationship between MBE performance and Passing rates.

Figure 1

Relationship between Bar Passage Rates and MBE Performance
On the Texas Bar Examination
2013 through 2017
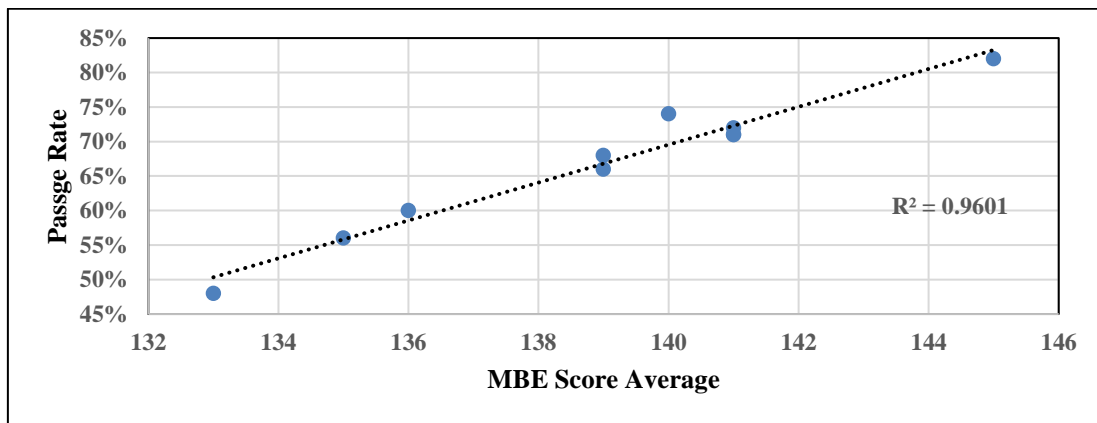
2. Performance of First-Time Takers and Repeaters

All U.S. Jurisdictions allow applicants to repeat the bar examination after a failed attempt. The rules for when an applicant can retake the examination and how many times they can retake vary from state to state. Texas is one of four states that limits the number of total attempts to five[7]. In their General Statistics reporting, the TBLE reports the examinee passage rates separately by first-time takers and all repeaters as a group. They do not provide descriptive statistics on the actual scores earned by those groups, nor do they disaggregate the statistics by the number of previous attempts.

In order to evaluate the size of the score differences between first-time takers and repeaters, as well as to determine if there are any patterns in performance between applicants retaking the exam more than twice, we calculated the passage rates and average Total Scale Scores by number of attempts. Table 6 presents these results.

Table 6

Bar Passage Rates and
Average Final Total Scale Scores
On the Texas Bar Examination
2013 through 2017

	First Time			2nd Attempt			3rd Attempt			4th or more		
	% Takers	Ave. Total	Pass Rate	% Takers	Ave. Total	Pass Rate	% Takers	Ave. Total	Pass Rate	% Takers	Ave. Total	Pass Rate
February												
2013	62%	716	81%	28%	680	67%	6%	663	51%	4%	670	59%
2014	64%	708	77%	25%	672	57%	5%	652	32%	5%	661	46%
2015	54%	695	69%	35%	670	54%	6%	660	41%	4%	641	30%
2016	48%	687	65%	39%	667	53%	7%	648	35%	6%	657	32%
2017	49%	683	61%	35%	652	38%	9%	648	31%	7%	647	32%
5-Year	55%	698	71%	33%	667	53%	7%	654	38%	5%	654	38%
July												
2013	90%	732	86%	5%	654	41%	4%	665	50%	2%	660	48%
2014	87%	711	77%	7%	648	35%	4%	647	33%	2%	647	27%
2015	84%	702	72%	7%	647	40%	6%	658	40%	3%	638	22%
2016	80%	715	78%	9%	654	46%	7%	653	39%	4%	653	36%
2017	79%	717	78%	9%	658	44%	8%	662	52%	4%	663	48%
5-Year	84%	716	78%	7%	653	42%	6%	657	43%	3%	653	37%

Data from Table 6 indicates that, similar to other jurisdictions, repeaters make up a significantly larger proportion of the February administration than July. Across the study period, test-takers attempting the exam three or more times comprised 12% of the entire testing group in February

---

[7] Texas does allow special appeals to retake after five failed attempts

and 9% in July. Those rates varied by year and have increased slightly in recent years as exam scores and passage rates among repeaters have decreased.

Over the five-year period, a steady decline can be observed in the average Total Scale Score of first-time takers sitting for the February administration (from 716 to 683; a change of 33 points). A smaller and less consistent decrease was seen for July first-time takers between 2013 and 2017 (from 732 to 717; a decline of 15 points). The changes in average scores were accompanied by decreases in the bar passage rates across the five years for this group (20% drop in February and 8% in July).

During this same period, applicants repeating the examination on February administrations also saw both their scores and passage rates decrease (28 points, 29% for first time repeaters; 25 points, 20% for applicants attempting a third time; and 23 points, 27% for applicants attempting a fourth or more). A very different pattern was observed for the July administrations. While average score performance varied between the years, examinees taking the exam for a second time saw both their average scores (654 to 658) and passage rates (41% to 44%) increase slightly from 2013 to 2017. A similar pattern occurred for July examinees attempting the examination for a third and fourth or more time.

For the July administrations (other than 2014 and 2015), we saw minimal difference in the average performance of applicants taking the exam two or more times. On the February administrations, the overall pass rates were identical for test takers sitting for the exam three or more times (38%), while the overall pass rate for those sitting a second time was 53%.

We suspect, as in other jurisdictions, that the applicants sitting for February administrations of the TBE are qualitatively different and represent a different group of applicants i.e., a disproportionate number of lawyers and foreign trained students, relative to those taking the July examination.

3. Correlations Among Test Sections

Each section of the TBE employs alternative methods to assess different domains of legal knowledge and skills. The knowledge and skill domains being assessed, however, are not completely independent of each other. The degree to which performance on one section relates to performance on another section can be measured by means of a correlation coefficient. This coefficient ranges in values from -1.00 to +1.00 with higher positive numbers reflecting a greater degree of overlap.[8] The size of a correlation coefficient can be impacted by several factors including the amount of variation in the respective scores, the commonality in skills & knowledge required on each measure, the similarity in format, and the reliability of each measure (see Section 6 for discussion of reliability).

Table 7 on the following page presents the inter-correlations of the different sections of the TBE for the years under study for both February and July administrations. We've also calculated the average correlation across the years.

With respect to the MBE, Table 7 reveals that, on average, correlations with other sections tend to be slightly higher in July than February. We believe that this is due to the larger variation in applicants' underlying proficiency in July than February. The moderate correlations observed between the MBE and Essay sections (averaging .59 in February and .64 in July), are in range of the relationship observed in most jurisdictions. The size of the MBE/Essay relationship is virtually identical to that observed between the MBE and the P&E section (averaging .59 in February and .63 in July). However, the fact that the Essay and P&E are only moderately correlated themselves (.65 and .67 on average in February and July, respectively), suggests that each of these examination sections are measuring a combination of both common and unique legal skills.

The lowest correlations in Table 7 are those between the MPT and the other TBE sections. The February and July average correlations of the MPT with the MBE are .29 and .36; with the Essay are .35 and .38, and with the P&E are .28 and .29. These low correlations are most likely due to the fact that the MPT receives a single holistic grade on only a six-point scale. By itself, a single MPT exercise lacks the reliability to demonstrate a higher degree of correlation with the other measures. An alternative interpretation is that the MPT measures a unique skill not captured by the other measures. However, the correlations between the MPT and Essay/MBE sections in other jurisdictions where multiple MPT questions are used, are found to be much higher. This suggests that the lack of relationship of the MPT is most likely a reliability issue associated with a single MPT item (see Section 6 for further discussion).

We observed very stable year-over-year, between-section relationships, with only minor variations. The larger variations involved the single MPT item, which lends further credence to the lowered-reliability interpretation.

---

[8] The correlation coefficient measures the consistency in rank ordering of individuals on two measures. A correlation of 1.00 between two variables would imply that individuals would be rank ordered exactly the same way on one measure as on the other. Interpretations of the size of correlations tend to vary, but generally, a .9 and above is considered very high; .7 to .9 is high; .5 to .7 is moderate; .3 to .5 is low; and 0 to .3 is indicative of little or no relationship.

Overall, the combination of stable relationships over time, and the moderate correlations between sections suggest that the TBE, through use of multiple measures and multiple formats, is consistently measuring a set of diverse legal content areas and skill domains in a consistent fashion.

Table 7

**Correlations Among Test Sections
Of the Texas Bar Examination
2013 through 2017**

	February					
	2013	**2014**	**2015**	**2016**	**2017**	**Ave.**
MBE						
Essay	.57	.58	.56	.61	.62	.59
P&E	.55	.58	.56	.64	.60	.59
MPT	.23	.24	.37	.33	.30	.29
Total	.86	.87	.87	.88	.89	.87
Essay						
P&E	.65	.64	.61	.68	.65	.65
MPT	.33	.34	.39	.38	.33	.35
Total	.73	.89	.87	.89	.90	.86
P&E						
MPT	.23	.24	.28	.34	.33	.28
Total	.74	.75	.71	.79	.76	.75
MPT						
Total	.42	.44	.52	.50	.46	.47
	July					
	2013	**2014**	**2015**	**2016**	**2017**	**Ave.**
MBE						
Essay	.61	.60	.62	.64	.71	.64
P&E	.61	.61	.65	.68	.62	.63
MPT	.37	.33	.37	.38	.33	.36
Total	.88	.88	.89	.90	.91	.89
Essay						
P&E	.66	.64	.68	.70	.65	.67
MPT	.42	.45	.34	.37	.32	.38
Total	.89	.89	.89	.90	.91	.90
P&E						
MPT	.23	.32	.30	.33	.25	.29
Total	.74	.75	.78	.80	.74	.76
MPT						
Total	.42	.53	.49	.51	.45	.48

5. Impact of The Re-grade Process

At the conclusion of the initial phase of grading, examinees either pass outright, fail outright or move into a second round of grading if their Total Scale Score falls between 669 and 674, inclusive. Examinees in the last group are eligible to have their Essay responses, P&E responses and their MPT responses re-evaluated to determine whether additional points can be added to their scores (points cannot be subtracted). Currently, there is no published information on the number of applicants that go into re-grade, the numbers that receive additional score points and the number of points received. Nor is the impact of the overall re-grade process on the final TBE passage rates documented.

We examined the extent and impact of re-grading by first determining which applicants went into re-grade over the five-year study period and then evaluating, in detail, what happened to them. This section presents the findings which emerged.

Applicants Going to Re-grade. Table 8 contains the counts of applicants going into re-grade on each examination administration along with the number getting extra score points by sections of the examination.   The final column on the right presents data on the mean number of essay questions for which the applicant was awarded extra points, along with the standard deviation for that mean, for each test administration during the five year study period.

Table 8
The Count of Examinees Going into Re-grade
And the Number Receiving Additional Score Points
By Section on the Texas Bar Examination
2013 through 2017

| | Number and Percentage* of Examinees Undergoing Re-grading | | Number and Percentage of Re-graded Examinees Receiving Additional Score Points | | | | | | | | | | No. of Essays Receiving Pts | |
| | | | PE Civil | | PE Criminal | | MPT | | Essay | | | | | |
	N	%*	N	%	N	%	N	%	N	%			Ave	Std
February														
2013	51	4.3%	2	4%	26	51%	12	24%	51	100%			6.5	1.7
2014	46	4.0%	42	91%	10	22%	4	9%	45	98%			3.5	1.4
2015	57	4.3%	57	100%	22	39%	4	7%	57	100%			4.7	1.5
2016	63	4.4%	28	44%	47	75%	8	13%	62	98%			3.3	1.5
2017	44	3.5%	36	82%	9	20%	6	14%	44	100%			3.7	1.6
5-Year	261	4.1%	165	63%	114	44%	34	13%	259	99%			4.3	1.9
July														
2013	87	2.9%	3	3%	54	62%	13	15%	87	100%			7.2	1.3
2014	87	3.0%	64	74%	26	30%	4	5%	87	100%			5.4	1.5
2015	119	4.0%	89	75%	119	100%	3	3%	118	99%			4.6	1.6
2016	97	3.3%	78	80%	22	23%	24	25%	95	98%			3.7	1.8
2017	78	2.6%	62	79%	27	35%	45	58%	76	97%			3.0	1.5
5-Year	468	3.1%	296	63%	248	53%	89	19%	463	99%			4.8	2.1

* Percentage of the total number of test takers

Results in Table 8 indicate that over the study period, an average of 4.1% of all applicants tested in February and 3.1% of applicants tested in July were eligible for re-grade, with slight variations between years.   In terms of which sections earned additional points, we observed significant differences. On the Essay section, nearly every applicant in re-grade gained extra score points on each administration, though the average number of questions garnering the extra points did vary. On the 2013 examinations, re-graded test-takers were awarded extra score points on approximately seven of the twelve essay questions, while on the subsequent examinations extra points were received on three to five questions.

On the MPT section, far fewer applicants received extra points, however. With the exception of July 2017 exam, where more than half of examinees in re-grade received extra points, only about one in ten applicants on the February examination and two in ten on the July examination received extra score points. And finally, with respect to the P&E tests, results varied widely. For example, in July 2015, all 119 applicants earned extra points, while only 23% earned points in July 2016. No consistent patterns could be observed in relation to this section of the exam.

Extra Points Earned in Re-grade. Because of the different weights applied to the respective test sections, extra score points earned on the Essay section will have a much larger impact on an applicant's final score than those earned on either the MPT or P&E sections. Table 9 provides a brief summary of the average number of points earned during the re-grade period over the five-year study period.

.

Table 9

**Average Number of
Score Points Earned in Re-grade
By Section on the Texas Bar Examination
2013 through 2017**

	Essay	MPT	P&E
February			
2013	5.3	1.6	0.8
2014	2.5	0.6	1.8
2015	3.5	0.9	3.2
2016	2.2	1.6	3.0
2017	2.8	1.7	1.6
5-Year	3.2	1.3	2.2
July			
2013	8.1	1.6	1.5
2014	4.3	0.6	1.4
2015	4.0	0.3	8.4
2016	4.3	3.7	2.1
2017	2.5	8.5	2.3
5-Year	4.6	2.7	3.5

Data in Table 9 highlights the fact that the number of additional scale score points earned on a given test section during re-grade varies significantly from year to year. Further, more points are likely to be earned on a July administration than a February one. This is somewhat puzzling given the fact that test-takers in the re-grade band, are in theory, of equal ability.[9] Differences between years on a given test section could be a function of the intricacies of scoring guides, the varying leniency standards of graders, or other extraneous factors.

Impact on Passage Rates. In the 2014 audit of the TBE, it was discovered that <u>all</u> test-takers who were in the re-grade range during 2013, went on to pass. At the time, it was unclear if this was a one-year phenomenon, or usual practice. Table 10 extends the original analysis for the five-year period.

<div align="center">

Table 10

**The Impact of the Re-grade Phase
On Examinee Pass/Fail Status
On the Texas Bar Examination
2013 through 2017**

</div>

	Total Applicants	Pass: Pre-Regrade		Fail: Pre-Regrade		Pass: In Regrade		Fail: In Regrade		Pass: Overall	
		N	%*	N	%	N	%	N	%	N	%
February											
2013	1,185	828	70%	306	26%	51	4.3%	0	0.0%	879	74%
2014	1,152	739	64%	367	32%	42	3.6%	4	0.3%	781	68%
2015	1,333	751	56%	525	39%	55	4.1%	2	0.2%	806	60%
2016	1,433	754	53%	616	43%	52	3.6%	11	0.8%	806	56%
2017	1,253	566	45%	643	51%	40	3.2%	4	0.3%	606	48%
5-Year	6,356	3,638	57%	2,457	39%	240	3.8%	21	0.3%	3,878	61%
July											
2013	3,023	2,387	79%	549	18%	87	2.9%	0	0.0%	2,474	82%
2014	2,929	2,009	69%	833	28%	82	2.8%	5	0.2%	2,091	71%
2015	2,987	1,866	62%	1,002	34%	119	4.0%	0	0.0%	1,985	66%
2016	2,975	2,014	68%	864	29%	84	2.8%	13	0.4%	2,098	71%
2017	2,959	2,053	69%	828	28%	71	2.4%	7	0.2%	2,124	72%
5-Year	14,873	10,329	69%	4,076	27%	443	3.0%	25	0.2%	10,772	72%

*Reported percentages are for the entire population of test-takers

Table 10 reveals that across all five years of the February administration, an additional 3.8% of all applicants received enough additional score points to pass, as compared to an additional 3.0% in July. Of the 729 test-takers who went to re-grade over this five-year period, only 46 failed to get sufficient additional points to pass. This group represented only 6% of all re-graded applicants, and .2% of all test-takers. Conversely, 683 out of the 729 (94%) test-takers in re-

---

[9] Statistical tests of significance revealed that the differences in earned re-grade points observed between both years and month of administration were significantly different (p<.001.)

grade between 2013 and 2017 passed. On 3 of the 10 administrations, no re-graded applicants failed. We can safely conclude that anyone who has written answers re-graded has an extremely high probability of passing the examination.

Given that the re-grade range extends from 669 to 674, this finding begs the question as to whether there is a point in that range where no one has failed in re-grade. We looked at the passage rates at each of the initial Total Scale Score points (i.e., 669, 670, etc.). We discovered that *every* test-taker entering re-grade with a score of either 673 or 674 during the 5-year period eventually passed. From there, 95% passed with initial scores of 671 or 672, 88% passed with an initial score of 670 and 82% passed with an initial score of 669. These statistics should be factored in if the TBLE considers alternative strategies to speeding up the grading process and score reporting.

6. Examination Reliability

Perhaps the most important measure of the psychometric quality of a test is its reliability. Reliability is a measure of the consistency or stability of scores. According to the National Council for Measurement, reliability is:

> "the characteristic of a set of test scores that relates to the amount of random error from the measurement process that might be embedded in the scores. Scores that are highly reliable are accurate, reproducible, and consistent from one testing occasion to another. That is, if the testing process were repeated with a group of test takers, essentially the same results would be obtained. Various kinds of reliability coefficients, with values ranging between 0.00 (much error) and 1.00 (no error), are usually used to indicate the amount of error in the scores."

Multiple factors impact the reliability of an examination including, the format(s), its length, test-taker instructions, the quality and consistency of subjective grader assessments, the conditions under which the test is administered, and underlying differences in the population taking the test.

On a high stakes test such as the bar examination, it is commonly agreed that reliability estimates above .85 should be achieved and estimates of .90 and above are preferable. In their 2014 audit of the TBE, Klein and Bolus reported overall TBE reliabilities of .89 and .86 for the 2013 February and July administrations. We replicated those calculations for the remaining years under study using (a) the methods of Cronbach (1951) to evaluate the reliability of the Essay and P&E sections, (b) the internal consistency reliabilities of the MBE, annually reported by the National Conference of Bar Examiners (February 2013 through July 2017) and (c) the computational methods of Rozeboom (1989) to estimate the weighted, linear composite of the Total Scale Score.[10]   The results of the calculations are presented in Table 11 on the following page.

With respect to the individual test sections, the 175-item MBE by itself has had reliabilities consistently over .90 since 2013, and reliability has steadily increased over that period. Reliability estimates on the 12-question Essay sections have remained fairly consistent, ranging from .76 to .82 on the February administrations and .79 to .82 on the July examinations. These are well within range of the internal consistencies reported by other states having a 12-question essay section (e.g., Ohio). The 40 combined items on the two P&E tests have internal consistency reliabilities averaging .83 in February and .81 in July.

The Total Score reliability estimates of the TBE are consistently strong and meet expected standards for a bar examination. On every administration, the Total Score reliability approached or exceeded the .90 level. This was due in major part to the high reliability of the MBE section and the 40% weight that it has been given in the overall score calculations. In the event that the TBLE opts to shorten the current examination to two days as is the practice in most states,

---

[10] As the MPT is only a single measurement, we do not have a direct method for estimating its reliability. Since the MPT is a 90-minute exam, we used a proxy of .50 which is the average of multiple configurations of three 30-minute essay questions. This is the same value used by Klein and Bolus in 2014.

consideration should be given to increasing the weight of the MBE if these same high levels of overall score reliability are to be achieved.

Table 11

**Estimates of the Score Reliability of
Section and Total Scores
On the Texas Bar Examination
2013 through 2017**

	MBE	Essay	P&E	Total Score
February				
2013	.90	.80	.84	.892
2014	.90	.82	.84	.898
2015	.90	.76	.84	.881
2016	.91	.77	.83	.892
2017	.92	.78	.79	.894
5-Yr Ave.	.91	.79	.83	.891
July				
2013	.90	.79	.82	.868
2014	.92	.79	.77	.900
2015	.92	.78	.83	.896
2016	.93	.79	.80	.903
2017	.93	.82	.81	.912
5-Yr Ave.	.92	.79	.81	.896

July 2016 P&E raw data was not available. It was
estimated reliability based on average of 4 other years

D. SUMMARY AND CONCLUSIONS

The results from this report indicate that TBE has functioned well over the five year period examined. Our replications of TBE calculated scores show that scoring procedures have been followed per published protocols. Applicant performance on the TBE has been comparable to all U.S. examination test-takers as evidenced by the similarities in mean MBE scores, while the overall distributions of both the individual sections and the over Total Scale Score have remained well-conditioned, contributing to the consistently exceptional reliability of test scores. Over the study period, examination reliability has approached or exceeded .90, achieving the target standard for a high stakes licensing examination.

In terms of passage rates, the Texas statistics are in line with those observed in the rest of the U.S. where success rates have also been on the decline. Among July first timers, the passage rate declined by 8% from 2013 to 2016, achieving a five-year low in 2015 (72%). The decrease on the February administrations was somewhat more pronounced (a 20% decrease; from 81% in 2013 to 61% in 2017). Additionally, as true for other jurisdictions, there was a steady drop off in passing rates relative to the number of previous attempts. In July, the passage rates were similar for those repeating a 2nd and 3rd time (42% vs 43%), with a slight drop for those repeating four or more times (37%). Overall, February's first-time repeaters performed slightly better than July repeaters (53% passing), while the passage rates of those repeating two or more times in February were consistently 38%.

In terms of the re-grade process, overall 4.1% and 3.1% of all examinees (729 over the five-year period) had their answers reevaluated in February and July, respectively. Of those examinees, 683 (94%) eventually passed. Fully 100% of the examinees with an initial score of 672 or 673 earned enough points to eventually pass, while on three of the 10 administrations under study, no re-graded examinee failed. Results further showed, that the majority of additional scale score points are earned on the Essay portion of the examination (relative to the MPT and P&E tests), and since it is the most heavily weighted constructed response section, it has had the major net impact on the outcome of this grading phase.

As the TBLE considers alternatives to its current examination structure, it should take into consideration the following key outcomes of this study:

First, the excellent reliability of the examination scores has been a function of the high levels of reliability of the MBE and the Essay sections. The Essay reliability is a function of the following factors: (a) its length (12 questions); (b) single-reader grading; and (c) effective us of the 25 point-per-question grading scale. If any of these factors were to be modified, the statistical reliability of that section, and the resulting reliability of the overall test could be affected.

Secondly, the current weighting structure gives equivalent weight to the MBE and Essay sections, and only ¼ of the weight to the remaining two parts of the exam. As the structure and/or configuration changes (e.g., decreasing the length of the Essay section or increasing the number of Performance Tests), there will most likely be a net effect in the overall consistency of

the Total Scale Scores. Consideration may need to be given to adjusting the weighting schema. This is why NCBE has adopted UBE scoring protocols where a higher weight is given to the MBE than either the Essay or MPT portions.

Finally, in the effort to improve score reporting for bar applicants, jurisdictions often consider modifying the grading process so as to release results faster. Adding graders or adjusting the re-grade process are sometimes considered as methods for achieving this objective. As these results have shown, however, Texas has achieved fairly strong written score reliability under its current configuration and single-reader per question grading procedures. The addition of more graders would potentially add more sources of measurement error and could subsequently reduce reliability.[11]

With respect to the re-grade process, the results from these analyses revealed that Texas' second read process, which generally takes at least a few weeks to complete, resulted in the passing of almost all applicants (though not necessarily 100% as shown in Klein & Bolus 2014 report). Consideration could be given to actually reducing the passing score (e.g., by two to three points on the current scale) and subsequently eliminating the re-grade phase altogether or reducing the re-grade range thereby reducing the number of examinees that would get re-graded. We understand that such a change could be met with objection from various sources, but study findings would support its soundness and rationale.

---

[11] Strong calibration procedures can minimize the effect of using multiple graders, but generally increases costs and administrative requirements

REFERENCES

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *The Standards for Educational and Psychological Testing (2014).*

Cronbach L. (1951). *"Coefficient alpha and the internal structure of tests".* Psychometrika. **16** (3): 297–334.

Hecht N. et al. for The Supreme Court of Texas (2016*). Order Establishing Task Force on the Texas Bar Examination.* Misc. Docket No 16-9104.

Klein P & Bolus R. (2005). *Analysis of July 2004 Texas Bar Exam Results by Gender and Racial/Ethnic Group*. A report prepared for the Texas Board of Law Examiners

Klein P & Bolus R. (2014). *Psychometric Audit of the Texas Bar Examinations Administered in 2013*. A report prepared for the Texas Board of Law Examiners

National Conference of Bar Examiners Website. (February 2013 through July 2017) Statistical Summaries. http://www.ncbex.org/statistics-and-research/statistics/

National Council for Measurement in Education Website http://www.ncme.org/ncme/NCME/Resource_Center/Glossary/NCME/Resource_Center/Glossary1.aspx?hkey=4bb87415-44dc-4088-9ed9-e8515326a061#anchorR

Rozeboom, W. (1989). *The Reliability of a Linear Composite of Nonequivalent Subtests*. Applied Psychological Management, Vol. 13, No. 3, pp 277-283