

# **PSYCHOMETRIC AUDIT OF THE TEXAS BAR EXAMINATIONS ADMINISTERED IN 2013**

Stephen Klein, Ph.D. and Roger Bolus, Ph.D.  
March 3, 2014

## **OVERVIEW**

We conducted our initial audit of the Texas Bar Exam in 2008. That audit and no doubt other factors led the Texas Board of Law Examiners to modify its bar exam procedures. The Texas Board has now asked us to review its current procedures and their effects, and to let it know whether its procedures are consistent with “best practice” for a large-scale high-stakes bar exam.

This report responds to that request. We begin by describing the exam’s four component tests, the methods it uses to compute total scores across those tests and make pass/fail decisions, the rationale for those procedures, and their effects on the exam’s statistical characteristics and quality. We extend this discussion to include a review of the exam’s reread policies and the effects and implications of those policies.

Next, we review the findings from our in-depth analyses of the “written” (as distinct from multiple choice) sections of the exam and the implications of those findings for improving test quality. We also discuss why the variation in passing rates on the Texas exam over the last 15 years (from a low of about 55% to a high of almost 85%) cannot be due to variation in the characteristics and difficulty of the questions asked or the graders employed or their grading practices. We can confidently rule out these factors because the procedures Texas adopted 15 years ago adjust the scores for possible differences in test difficulty over time. In short, variations in question difficulty and/or reader leniency (other than in reread) cannot affect Texas’ passing rates.

We conclude this audit with a summary of our major findings and their implications; and our recommendations for next steps.

## **EXAM SECTIONS**

The Texas Bar Exam is administered in February and July by the Texas Board of Law Examiners. Each administration takes 2.5 days. The four sections of this exam and their scoring rules are discussed below:

**Multistate Bar Exam (MBE).** The MBE is a 6-hour, 190-item multiple-choice test of six areas of the law that are typically covered in the first two years of law school.<sup>1</sup> The MBE is now used by all but one state and is constructed and scored by the National Conference of Bar Examiners (NCBE). NCBE converts the “raw” scores on this test (i.e., the number of questions the candidate answered correctly) to “scale” scores using standard psychometric methods. This conversion calibrates the results obtained by current test takers to those obtained by previous test takers on a representative sample of the items that were included in both the current test and one or more previous versions of that test.

For example, if current test takers do just as well on the repeated items as previous test takers performed on those same items, then the current group should earn about the same total score on them as the previous group. If they don’t, then the current group’s scores are adjusted up or down to make them equivalent. Because of this step, a given MBE scale score indicates about the same level of proficiency regardless of the particular administration of the exam on which that score was earned. As a result of this and other features of the MBE almost all states, including Texas, convert their non-MBE scores to the MBE’s scale of measurement to compute a candidate’s total score and pass/fail status.

**Multistate Performance Test (MPT).** This 90-minute test is designed to assess an examinee’s ability to complete a task that a beginning lawyer should be able to perform using realistic case materials. NCBE consultants draft the MPT tasks and the “library” of case materials candidates may use in preparing their answers. Almost 75% of the states administer one or two MPT items but they differ in how much weight they give them and how closely they adhere to the NCBE’s scoring guide. Texas assigns each answer randomly to one of two trained readers for grading.

**Procedures and Evidence Examination (P&E).** This 90-minute test consists of 20 short-answer Criminal Law questions and 20 short-answer Civil Law questions. All the questions are constructed by Texas. Each answer is graded on a 5-point scale following a detailed analytic scoring guide. Total P&E raw scores are divided by 2 so that the maximum Civil and Criminal raw scores are each worth 100 points. One reader grades all the answers to the Civil Law questions and a different reader grades all the answers to the Criminal Law questions. Texas is the only state that administers this type of test.

**Essay Test.** This 12-question test is administered in two 3-hour blocks with 6 questions per block. Texas drafts all the questions and grades the answers. One reader grades all the answers to a given question. Readers assign grades on a 1 to 25 point scale. Applicants may hand write or use their laptops to key enter their answers. One Board member supervises the grading for each of six pairs of essay questions. New sets of Essay (and P&E and MPT questions) are constructed for each administration.

---

<sup>1</sup> The six areas are Contracts, Constitutional Law, Criminal Law and Procedure, Evidence, Real Property, and Torts. Federal Jurisdiction is being added in February 2014. Applicants also answer 10 items that are being pretested for future use. The applicants’ answers to these 10 items do not affect their MBE scores.

## **COMPUTING CONVERTED AND TOTAL SCALE SCORES**

The “raw” score on a multiple choice test is the number of questions answered correctly. However, because of differences in test difficulty, it may be more difficult to earn a given raw score on one test than on another test even if both tests covered the same content and skill and asked the same number of questions. Moreover, when raw scores from different tests are summed to obtain a total “composite” score, the test with the largest score spread of raw scores will carry the most weight in determining a student’s relative standing on the total exam. The “standard deviation” (sd) is the typical index of score spread. A distribution with a very narrow spread, such as one that is shaped like the Washington Monument, has a much smaller sd than a distribution that looks like a pitcher’s mound.

Thus, to assure that the scores on different sections of a test are weighted and valued appropriately, their scores are transformed to a common scale of measurement, much like converting the currencies of different countries into Euros.

The Texas scaling process begins by converting the reader assigned raw scores on each of the 12 essay questions to a score distribution that has a mean of 100 and a sd of 10 so that all the Essay questions are on the same scale of measurement and carry the same weight in determining an applicant’s total Essay score. Next, the sum of each applicant’s 12 converted Essay question scores is computed and the distribution of these sums is converted to a distribution that has a mean and sd that are equivalent to the MBE’s mean and sd among all Texas applicants taking the same exams. The MPT, P&E-Civil, and P&E-Criminal scores are also converted to score distributions that have the same mean and sd as their MBE scale scores in Texas. Finally, the formula below is used to compute weighted total scores so that the MBE and Essay sections each carry 40%, the MPT carries 10%, and the P&E Civil and Criminal each carry 5%).<sup>2</sup>

Total TX scale score = (P&E Scale)/2 + (MPT Scale)/2 + 2(MBE Scale) + 2(Essay Scale)

The Board provides an abbreviated description of the exam, its sections and scoring procedures to the public on its website: ([http://www.ble.state.tx.us/exam\\_info/Grading%20Explanation%20as%20of%201-11-08\\_pdf.pdf](http://www.ble.state.tx.us/exam_info/Grading%20Explanation%20as%20of%201-11-08_pdf.pdf))

We applied the above procedures to the raw scores of the examinees who took the February or July 2013 examination. Our calculations arrived at the exact same converted, section scale and total scale scores as calculated by the Texas Board, thereby confirming that the scoring procedures are being applied correctly and as publicized.

---

<sup>2</sup> Each scale score (except the MBE’s) is rounded to two decimal points before being put into the total score formula. MBE scale scores are rounded to one place because NCBE only rounds them to one place. As a result of the scaling process described above, one MBE scale score point usually equals about 5 Texas scale score points.

Table 1 shows the mean, standard deviation and score reliability on each section (except the MBE) before scaling and weighting. MBE scores are scaled by NCBE before Texas receives them. “Composite” score reliabilities (i.e., the reliability of the sum of two or more scores (such as the sum of the MBE and Essay scale scores) were computed using Rozeboon’s method (*Applied Psychological Methods*, 1989, Vol.13, 277-283).

Table 1  
Raw Score Means, Standard Deviations, And Score Reliability  
On Each Section Before Scaling.

| Score          | February 2013 |       |       | July 2013 |       |       |
|----------------|---------------|-------|-------|-----------|-------|-------|
|                | Mean          | SD    | Alpha | Mean      | SD    | Alpha |
| MBE            | 140.1         | 13.51 | .90   | 144.8     | 14.97 | .90   |
| Essay          | 1201          | 67.23 | .80   | 1201      | 66.07 | .79   |
| MPT            | 3.22          | 1.17  |       | 3.22      | 1.17  |       |
| P&E – Criminal | 64.66         | 14.24 | .72   | 68.47     | 13.24 | .66   |
| P&E – Civil    | 64.98         | 13.78 | .77   | 64.19     | 11.47 | .76   |
| Total          |               |       | .89   |           |       | .86   |

Shaded cells designate values that cannot be computed from the available data.

### **FACTORS AFFECTING AND NOT AFFECTING THE PASSING RATE**

One important consequence of these rules is that the Board cannot affect the exam’s difficulty or passing rate by asking harder or easier questions. The readers also cannot affect the exam’s difficulty or passing rate by grading more or less leniently. Test difficulty and passing rate (on first read) are all governed by the mean and sd of the MBE scores. The Texas Board and its readers have no control or influence over those scores. However, how much weight the Board places on the different sections of the exam and how they grade in reread can affect how many and which applicants pass.

Figure 1—which has one plotted point for each exam analyzed—shows that an exam’s mean MBE scale score (on the x-axis) explains virtually all (98%) of the variation in Texas passing rates (on the y-axis). Hence, the only way the Board can affect the passing rate is to adjust the total score required for passing or be more or less liberal in the scores they assign to the 4% of the candidates whose answers are reread. In short, decisions as to which and how many applicants pass are largely out of their hands.

### **REREREAD AND PASS/FAIL RULES AND THEIR EFFECTS**

Applicants with a total scale score of 675 or higher pass while those with total scores below 669 fail. All other applicants have all 12 of their essay answers, their MPT answer, and all of their P&E answers read again, but by the Board member who supervised the initial reader on the respective questions.

The re-readers may add to but not subtract from the score the initial reader assigned to an answer. Many states, including Texas, set their passing scores at a point that is equivalent to a 135 on the MBE scale or at a score very close to that (see Appendix A). However, there is very little if any agreement among states in their reread policies.

Texas passed every applicant in its reread zone on the February and July 2013 bar exams (see Table 2). Texas also passed almost every applicant in reread on its other recent bar exams. In short, Texas could eliminate reread entirely if it just passed everyone in the reread zone (and announce the passing score in Texas was 70 on the Texas scale (and 134 on the MBE scale).

Table 2  
Number And Percentage Of Applicants In Each Pass/Fail Category By Exam

| Pass/Fail Category   | February 2013 |         | July 2013 |         |
|----------------------|---------------|---------|-----------|---------|
|                      | Number        | Percent | Number    | Percent |
| Pass on initial read | 827           | 69.8    | 2,387     | 79.0    |
| Pass after reread    | 52            | 4.4     | 87        | 2.9     |
| Fail after reread    | 0             | 0       | 0         | 0       |
| Fail before reread   | 306           | 25.8    | 549       | 18.2    |
| Total                | 1,185         | 100.0   | 3,023     | 100.0   |

It is not clear why every (or almost every) applicant in reread passed. It could be due to: (a) the re-readers being more liberal in their grading than the initial readers, (b) the re-readers being influenced by the grades assigned by the initial readers, (c) some function of the rounding rules, or (d) some combination of these or other factors. The only way to find the cause is to have a large set of answers graded again but by readers who are “blinded” to the scores assigned by the first batch of readers.

## CONCLUSIONS & RECOMMENDATIONS

1. The Texas results are perfectly consistent with its published scoring procedures. The current software for creating converted and scale scores is operating correctly as are the calculations of total scores and assignment of pass-fail status.
2. The February and July 2013 total score reliability coefficients (.89 and .86, respectively), were among the highest we have seen for a bar exam.
3. To support continuous improvement and monitoring of exam quality, we suggest that Texas routinely perform pre and post-examination analyses. These analyses would include performing item analyses to check for possible problems with scoring rubrics. (The electronic files we gave Texas include item analyses of the P&E and Essay tests that would facilitate this type of monitoring). We also suggest running a check on an inter-reader agreement with a sample of MPT answers.

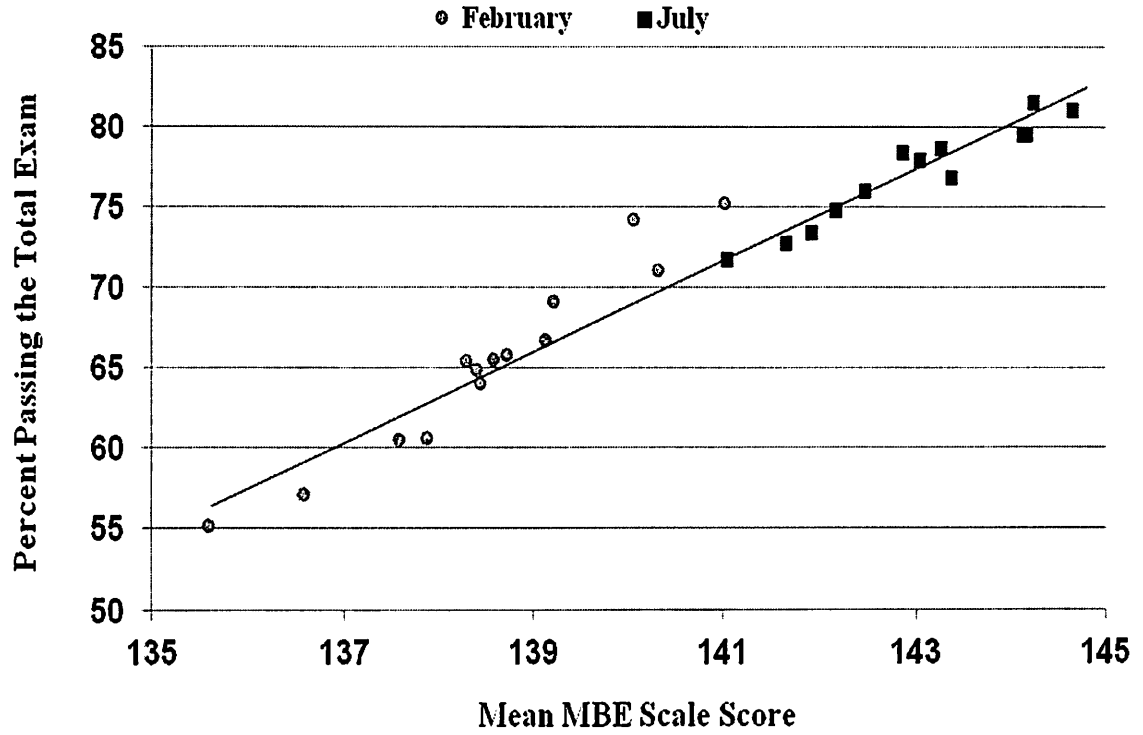
4. On the last several bar exams, all or virtually all of the applicants who had their answers reread passed. This finding suggests that none of the applicants who had initial total scores just below the one required to get them into reread would have failed if they had (1) gone to reread and (2) the re-readers were not aware of their pre-reread scores. In other words, the test was a perfect indicator of who should pass and fail. We doubt that any bar exam is that precise. To find out, one need only run a “double-blind” study in which a representative sample of answers are read again by graders who do not know the scores that were assigned by the initial readers and then assess how closely the scores assigned by the re-readers correspond to those assigned by the initial readers. How often such a study should be run would depend on the results obtained.

Our concerns about potential reader calibration problems are more than offset by Texas having higher than normal total score reliability, giving the MBE close to 50% of the weight in computing using total scores, using a “compensatory” rather than a “conjunctive” pass/fail rule, giving a 2.5-day (instead of the typical 2-day) exam, and using multiple question types (multiple choice, short answer, performance, and essay) to improve validity. All of these features should be retained.

5. Finally, the re-grading decisions that led to everyone in reread passing implies that everyone who fell below 670 on the initial grading deserves to fail and everyone who fell at or above 671 deserves to pass. In short, the test results are assumed to be a perfect discriminator between who should pass and fail. The “double-blind” grading study described above would provide a more realistic assessment of reader agreement; i.e., a study in which re-graders evaluate answers without knowing whether they agree with the grades assigned on the first reading or what effect re-grading would have on an applicant’s pass/fail status. We strongly recommend doing such a study in connection with at least one July and one February exam.

**Figure 1**

**Percent passing the Texas Bar Exam between February 2000 and July 2013  
as a function of the Mean MBE score in Texas on the exam date.**



NOTE: Each data point is a Bar Exam Administration Date.

EB96-1 01/2014

## Appendix A

### Official Reported Passing Score by State as of 2013

|     |   |
|-----|---|
| 145 | DE  |
| 144 | CA  |
| 143 |   |
| 142 | OR  |
| 141 |   |
| 140 | AK ID (NV) VA   |
| 139 |   |
| 138 | CO ME NC (RI)   |
| 137 | AZ  |
| 136 | FL PA   |
| 135 | AR GA MA MD MI MT NE NH OH OK TN TX UT (VT) WV WA<br>(WY) |
| 134 | HI TN   |
| 133 | DC IA KS NJ NY  |
| 132 | CT IL IN (KY) MS  |
| 131 |   |
| 130 | MN MO ND NM SD  |
| 129 | WI  |
| 128 | AL  |
| 127 |   |
| 126 |   |
| 125 | (SC)  |

Source: Chart 9: Grading and Scoring in National Conference of Bar Examiners Comprehensive Guide to Bar Admission Requirements, 2013, Reported Passing Scores.

Louisiana does not use the MBE and Wisconsin does not require graduates of ABA accredited in-state law schools to pass a bar exam. The difficulty of passing a state's exam is a function of its passing score and other rules, such as the number of times an applicant may repeat the exam and the degree of subject matter overlap between the MBE and other portions of the exam. States in parenthesis use conjunctive pass/fail rules as follows:

- NV  $\geq 140$  mean total scale and  $\geq 75$  raw on  $\geq 3$  of 9 essays.
- KY  $\geq 132$  on MBE and mean raw  $\geq 75$  on essay section.
- RI  $\geq 130$  on MBE and total scale score  $\geq 276$
- WY  $\geq 130$  on MBE and essay raw score  $\geq 70$  on 6 of 10 essays.
- VA  $\geq 130$  on both the MBE and total essay scale.
- SC  $\geq 125$  on MBE and  $\geq 70$  raw on all 6 essay sections.