



TOURO UNIVERSITY
JACOB D. FUCHSBERG LAW CENTER
Where Knowledge and Values Meet

Digital Commons @ Touro Law
Center

Scholarly Works

Faculty Scholarship

2016

UBE-Shopping: An Unintended Consequence of Portability?

Suzanne Darrow-Kleinhaus

Touro Law Center, sdarrow-kleinhaus@tourolaw.edu

Follow this and additional works at: <https://digitalcommons.tourolaw.edu/scholarlyworks>



Part of the [Legal Education Commons](#)

Recommended Citation

88 N.Y.ST. B.J. 46 (July/August, 2016)

This Article is brought to you for free and open access by the Faculty Scholarship at Digital Commons @ Touro Law Center. It has been accepted for inclusion in Scholarly Works by an authorized administrator of Digital Commons @ Touro Law Center. For more information, please contact lross@tourolaw.edu.

UBE-Shopping: An Unintended Consequence of Portability?

Suzanne Darrow-Kleinhaus¹

Introduction

Last year at this time, New York was engaged in determining whether to adopt the Uniform Bar Examination (UBE); this year, our May 2016 graduates are getting ready to take it in July.

Getting our students ready for the UBE, however, may require more than just learning the law; it also means learning in which jurisdiction you should take it. While there is not much that is new about the UBE's individual components — the Multistate Essay Examination (MEE), the Multistate Performance Test (MPT) and the Multistate Bar Examination (MBE)² — what is new is that where you take the UBE may make the

¹ Professor Darrow-Kleinhaus is the Director of Academic Development and Bar Programs at Touro College, Jacob D. Fuchsberg Law Center. In addition to books on developing skills for law school learning and the bar exam, including *MASTERING THE LAW SCHOOL EXAM*, *THE BAR EXAM IN A NUTSHELL*, *ACING THE BAR EXAM*, and *THE NEW YORK BAR EXAM BY THE ISSUE*, she has written law review articles in this area, *A Response to the Society of American Law Teachers' Statement on the Bar Exam and Incorporating Bar Pass Strategies into Routine Teachings Practices*. She has also published in the areas of contract law, labor and employment law, the Fourth Amendment Exclusionary Rule, and federal preemption.

Dr. Nancy Johnson, Ph.D., J.D.: A special acknowledgement is due to Dr. Johnson for her generosity of time and knowledge. Her expertise in statistical analysis and the bar exam were essential to the writing of this paper. It could not have been written without her guidance and contributions. Dr. Johnson is a California attorney and licensed psychologist in clinical psychology. She is the author or co-author of over 150 peer-reviewed publications and papers, including more than 20 law study guidebooks and an interdisciplinary book that critically reviewed the 1998 empirical literature on domestic violence. In conjunction with Dennis P. Saccuzzo, she has lectured extensively in law and has taught a variety of supplemental and full Bar programs for first time and repeating Bar candidates. She is co-founder of Applications of Psychology to Law, Inc., a corporation devoted to the application of the psychological sciences to the study of law.

² *UBE: Uniform Bar Examination*, NATIONAL CONFERENCE OF BAR EXAMINERS, <http://www.ncbex.org/exams/ube/> (last visited Mar. 24, 2016). The National Conference of Bar Examiners (NCBE) develops and sells these three test instruments to jurisdictions. The MBE is a multiple-choice exam with 200 questions testing examinees' knowledge of Civil Procedure, Constitutional Law, Contracts and UCC Article 2, Criminal Law and Procedure, Evidence, Real Property, and Torts. The MEE includes essay questions covering these MBE subject and five additional areas. The MPT consists of two performance tasks where examinees complete "lawyerly"

difference between passing and failing. This is possible because of the convergence of bar exam test practices of “portability,” “relative grading,” and “scaling” of scores.

By adopting the UBE, jurisdictions agree to weight the MEE at 30%, the MPT at 20%, and the MBE at 50% in determining an examinee’s score. As a result, the UBE, as currently administered,

- leads to the situation where the same skill level could result in different UBE scores depending on where the candidate takes the exam,
- fails to ensure that the scores used to grant licensure in a UBE jurisdiction are sufficiently reliable for high-stakes testing when it is possible to achieve different outcomes on the same test by the same candidate if taken in different UBE jurisdictions,
- results in a “portable” score but not an “accurate” one because the written score — 50% of the total — depends on the strength of the applicant pool in the jurisdiction where the candidate wrote the exam,
- presents a candidate with the opportunity to “UBE shop” and “game the system” by taking the UBE in a jurisdiction where the same essay and MPT performance would result in a higher score and then transferring that inflated score for admission in a “harder” jurisdiction,
- makes it possible for a candidate to file a discrimination lawsuit challenging his or her UBE results.

The possibility that one can “game the test” makes a UBE score inherently unreliable.

The National Conference of Bar Examiners (NCBE), the entity that produces the MEE, MPT, MBE, and MPRE components of the bar exam, claims that the UBE’s benefit to the public is that it “will provide more consistency in the requirements for bar admission across the country” and more consistency will, in turn, “make the bar

admission process more understandable to members of the public.”³ Neither claim, however, survives scrutiny.

NCBE claims that the UBE will provide more consistency in bar admission requirements.

NCBE asserts that the UBE provides more consistency in bar admission requirements than non-UBE jurisdictions because the UBE weights its components equally and scores them uniformly as opposed to differences among non-UBE jurisdictions in their grading and scoring.⁴ NCBE further claims that the UBE provides the consistency essential for comparisons between jurisdictions of examinees’ competency because all UBE examinees “will be taking exactly the same exam and receiving scores that will have the same meaning across the country.”⁵ The scores have the same meaning because the UBE is “uniformly administered, graded, and scored”⁶ by the jurisdictions that adopt it. Consequently, although UBE jurisdictions may set differing cut scores for admission, what remains “consistent” is the assurance that a

³ Susan M. Case, *The Testing Column, The Uniform Bar Examination: What’s In It For Me?* THE BAR EXAMINER, Feb. 2010, at 50, 52 [hereinafter Case, *What’s In It For Me?*] http://www.ncbex.org/pdfviewer/?file=%2Fassets%2Fmedia_files%2FBar-Examiner%2Farticles%2F2010%2F790110_TestingColumn.pdf.

⁴ See NATIONAL CONFERENCE OF BAR EXAMINERS, COMPREHENSIVE GUIDE TO BAR ADMISSION REQUIREMENTS 40 (2016), (Chart 9: Grading and Scoring, While the UBE weights the MBE at 50% and the written portion, MEE and MPT, at 50%, Chart 9 shows the range about jurisdictions: MBE weights range from 33 to 50 percent, the MEE and/or local essay exam from 25 to 60 percent, and the MPT or local performance test from 8.7 to 26 percent.) <http://www.ncbex.org/pubs/bar-admissions-guide/2016/index.html#p=1>.

⁵ Case, *What’s In It For Me?*, *supra* note 3, at 51. See also *UBE Score Portability*, NATIONAL CONFERENCE OF BAR EXAMINERS, <http://www.ncbex.org/exams/ube/score-portability/> (last visited Feb. 26, 2016). NCBE advises jurisdictions that because every UBE jurisdiction uses the same essay questions, the same performance tasks, and the same grading guidelines, as long as the candidate sits for all portions of the UBE in the same UBE jurisdiction and in the same administration, a portable UBE score is earned that can then be transferred to other states that have joined the UBE network.

⁶ *Jurisdictions That Have Adopted the UBE*, NATIONAL CONFERENCE OF BAR EXAMINERS, <http://www.ncbex.org/exams/ube/> (last visited Feb. 26, 2016).

UBE score represents an examinee's fitness for the practice of law within the UBE roster of jurisdictions.

The question, however, is whether the scoring process followed by UBE jurisdictions achieves this level of reliability. According to NCBE, “[h]igh reliability is essential [in high-stakes licensing tests] to ensure that the pass/fail status of examinees would not flip-flop from one administration to the next or if different questions were asked, if different graders were grading the papers, or if the examinees were testing with a more or less able group of examinees.”⁷ Accordingly, “[j]urisdictions that scale the essays to the MBE scores for their jurisdiction, that weight the MBE at least 50%, and that make the pass/fail decision on the total score are assured of a sufficiently high reliability and high decision consistency.”⁸

However, equal weighting of components and the same exam components does not assure that resulting scores are uniform so as to “have the same meaning.” An example shows why “uniform” scores may not “have the same meaning” and therefore may not be sufficiently reliable for high-stakes testing.

We begin with the mean MBE scores from the July 2015 bar exam.⁹ The following are those scores that are available from jurisdictions that publish their state's mean MBE score, although not the standard deviation (“s.d.”):

⁷ Susan M. Case, *The Testing Column, Quality Control For Developing and Grading Written Bar Exam Components*, THE BAR EXAMINER, June 2013, at 34, 36 [hereinafter Case, *Quality Control for Developing and Grading Written Bar Components*] http://ncbex.org/assets/media_files/Bar-Examiner/articles/2013/820213Testing-Column.pdf.

⁸ *Id.*

⁹ E-mail from Nancy E. Johnson to Suzanne Darrow-Kleinhaus, Professor of Law and Director of Academic Development and Bar Programs, Touro Law Center (Feb. 15, 2016, 2:49 p.m. EST) (on file with author).

July 2015 MBE MEAN FOR SELECTED JURISDICTIONS¹⁰

Jurisdiction	MBE Mean
California	142.4
Pennsylvania	142.2
Georgia	140.2
National	139.9
Tennessee	139.8

The largest difference in mean MBE score among these jurisdictions is 2.6.

Now consider NCBE's published national mean MBE score in 2014, as well as their standard deviation:

2014 National Mean and Standard Deviation of MBE SCORES¹¹

Date	MBE Nat. Mean	s.d.
July 2014	141.5	16
Feb. 2014	138.0	15.3

NCBE claims that their standardization process of equating makes it so that an MBE score of 140 in July has the same meaning as a 140 in February; the difference in mean from July to February is because the February candidates are weaker.¹² A July

¹⁰ The information in the table entitled *July 2015 MBE Mean for Selected Jurisdictions* was collected from multiple web sites. See Derek T. Muller, *California Bar Exam Takers Are Far More Able Than Others Nationwide But Fail At Much Higher Rates*, EXCESS OF DEMOCRACY, <http://excessofdemocracy.com/blog/2015/11/california-bar-exam-takers-are-far-more-able-than-others-nationwide-but-fail-at-much-higher-rates>, (last visited Mar. 28, 2016) (California Mean of 142.4); *July 2015 Pennsylvania Bar Examination Statistics*, PENNSYLVANIA BOARD OF LAW EXAMINERS, <http://www.pabarexam.org/pdf/statistics/july/j2015.pdf> (last visited Mar. 28, 2016) (Pennsylvania Mean of 142.4); *Georgia Bar Examination Statistics*, SUPREME COURT OF GEORGIA OFFICE OF BAR ADMISSIONS, <https://www.gabaradmissions.org/georgia-bar-examination-statistics#0715> (last visited Mar. 28, 2016) (Georgia Mean of 140.2); National Conference of Bar Examiners, *2015 Statistics*, THE BAR EXAMINER, Mar. 2016, at 14, 44 (Table entitled: 2015 MPRE National Summary Statistics Based on Scaled Scores) http://www.ncbex.org/pdfviewer/?file=%2Fassets%2Fmedia_files%2FBar-Examiner%2Fissues%2FBE-March2016-Abridged.pdf; *Statistics of the Tennessee Bar Exam 2013, 2014 and 2015*, TENNESSEE BAR EXAM (Oct. 25, 2015), <http://tennesseebarexam.blogspot.com/> (last visited March 28, 2016) (Tennessee Mean of 139.8).

¹¹ National Conference of Bar Examiners, *2014 Statistics*, THE BAR EXAMINER, Mar. 2015, at 8, 34, (Table entitled: 2014 MBE National Summary Statistics based on Scaled Scores) <http://www.ncbex.org/pdfviewer/?file=%2Fdmsdocument%2F164>.

¹² Susan M. Case, *The Testing Column, Demystifying Scaling To the MBE: How'd You Do That?*, THE BAR EXAMINER, May 2005, at 46 [hereinafter Case, *Demystifying Scaling To the MBE*],

candidate who is relatively weak on the MBE but better on the written would do better by taking the bar in February, when her essay score would be even higher because of the comparative grading.

By this logic, then, the candidates in July 2015 in Tennessee (mean MBE score of 139.8) are weaker than the candidates in California (mean MBE score 142.2). But the UBE scales a written score to the candidates in the jurisdiction where the UBE was taken. So because of the way the UBE is scaled, that same July candidate in a UBE jurisdiction could now achieve a similar result by picking a jurisdiction more like the February national pool where the mean MBE score is 138 and simply transferring that score to her own preferred jurisdiction (forum shopping, rather than deferring until February). But to do that, she would have to choose wisely so that her written performance is comparatively higher enough, and we don't have the information to make that choice.¹³

Consider an example:

Written scaled = (candidate written score s.d.)(statewide MBE s.d.) + (statewide mean MBE score)

Suppose our candidate scores 124 on the MBE: she would need 156 on the written to total 280, the UBE passing score in that jurisdiction. In her home jurisdiction (like the July national mean), the MBE mean is 141.5 and the s.d. is 16, so she would need to be 0.91 s.d. above the mean in her written performance (82nd percentile).

But now choose a jurisdiction where the pool looks more like a February pool, with a mean of 138 and s.d. of 15.3. Now she would need to be 1.2 s.d. above the mean (89th percentile in that weaker jurisdiction) to reach a 280.

www.ncbex.org/pdfviewer/?file=%2Fassets%2Fmedia_files%2FBar-Examiner%2Farticles%2F2005%2F740205_testing.pdf.

¹³ E-mail from Nancy E. Johnson to Suzanne Darrow-Kleinhaus, Professor of Law and Director of Academic Development and Bar Programs, Touro Law Center (Feb. 15, 2016, 2:49 p.m. EST) (on file with author).

It seems likely that our candidate's strong written performance in her jurisdiction would rank her even higher in a weaker jurisdiction, so it would work to forum shop, but once again, we do not have the necessary statistics to test the hypothesis. Testing requires knowing the MBE mean and the standard deviation from that mean for that jurisdiction because the essays and performance test raw scores are scaled using that number.

Nonetheless, it can be inferred that achieving a different numerical score for the exact same performance is possible depending on where the candidate wrote the exam because of "relative grading." Relative grading or "rank-ordering" occurs when graders make grading distinctions among papers where the "top grade does not necessary indicate an excellent paper; it just indicates a paper that is better than the other papers."¹⁴ For the UBE, this means that the examinee's written portion — the MEE and MPT — is scored "relative" to the other examinees' answers in that jurisdiction. These "ranked" scores are then scaled to the MBE.¹⁵

Returning to our example, if we apply this process to scoring our candidate's written bar exam components, she can "appear" better and therefore be "ranked" higher when in the company of one group as opposed to another. ***NCBE has acknowledged this situation: it has been shown that "an essay of average proficiency will be***

¹⁴ Case, *Quality Control for Developing and Grading Written Bar Exam Components*, *supra* note 7, at 36.

¹⁵ Judith A. Gundersen, *The Testing Column, Essay Grading Fundamentals*, THE BAR EXAMINER, March 2015, at 54, http://www.ncbex.org/assets/media_files/Bar-Examiner/articles/2015/840115-abridged.pdf. In rank-ordering, some papers "should get high scores, some average scores, and some lower scores, regardless of what score scale a jurisdiction uses (1 – 5, 1 – 6, 1 – 10, etc.), and regardless of whether, taken as a whole, papers are strong or weak. What matters is rank-ordering among papers — relative grading." For example, assuming a jurisdiction uses a 1 – 6 scale, a "1" paper is a very poor answer relative to the other answers in the jurisdiction and a "6" paper is an excellent answer relative to the other answers in the jurisdiction. However, this does not necessarily indicate that the top grade — the "6" paper — is an excellent paper; "it just indicates a paper that is better than the other papers."

graded lower if it appears in a pool of excellent essays than if it appears in a pool of poor essays. Context matters.”¹⁶ Finally, when this “ranked” score is then scaled to the MBE score for that group, she may end up with a higher UBE score than she would otherwise receive. Thus, while the score is “portable,” it is not accurate because the written score — 50 percent of the total — depends on the strength of the applicant pool in the jurisdiction where she wrote the exam.

The size of the applicant pool would also play a role, especially if that affects the standard deviation of the MBE distribution in that jurisdiction.¹⁷ This requires understanding how essay scores are scaled to the MBE. According to Dr. Susan Case, former Director of Testing for the National Conference of Bar Examiners, “[s]caling the essays to the MBE is an essential step in ensuring that scores have a consistent meaning over time. When essay scores are not scaled to the MBE, they tend to remain about the same: for example, it is common for the average raw July essay score to be similar to the average February score even if the July examinees are known to be more knowledgeable than the February examinees. Using raw essay scores rather than scaled essay scores tends to provide an unintended advantage to some examinees and an unintended disadvantage to others.”¹⁸

What Dr. Case is telling us is that essay graders engage in relative grading so that the top performers in a group get the same top scores as those in a prior group,

¹⁶ Susan M. Case, *The Testing Column, Frequently Asked Questions About Scaling Written Test Scores to the MBE*, THE BAR EXAMINER, Nov. 2006, at 43, http://www.ncbex.org/pdfviewer/?file=%2Fassets%2Fmedia_files%2FBar-Examiner%2Farticles%2F2006%2F750406_Testing.pdf.

¹⁷ E-mail from Nancy E. Johnson to Suzanne Darrow-Kleinhaus, Professor of Law and Director of Academic Development and Bar Programs, Touro Law Center (Feb.14, 2016, 11:26 a.m. EST) (on file with author).

¹⁸ Case, *Demystifying Scaling to the MBE*, *supra* note 12, at 46. Dr. Case was the Director of Testing until Nov. 1, 2013.

regardless of whether the pool is less competent than a prior pool. Scaling to the MBE is supposed to provide a consistent meaning over time because the national distribution of the MBE is equated across time and the raw scores across the country presumably approximate a normal distribution. But with the UBE, however, essays are not scaled to a national distribution (which we concede has been scaled across time). Instead, they are scaled to that jurisdiction's MBE distribution by forcing them to have the mean and standard deviation as that of the MBE distribution for that jurisdiction. In other words, the same skill level on the essays and MPT would get a different score in different jurisdictions, depending not only on the relative written skill of the jurisdiction's candidates, but also the relative MBE skill. This can have a significant impact on individual scores, especially in smaller jurisdictions.

Using the same method of scaling that NCBE¹⁹ uses, let's see what would happen with a hypothetical candidate. According to Dr. Nancy Johnson, assuming we have a candidate who scores 125 on the MBE when the national mean is 140 and the standard deviation is 15 (so this candidate is 1 s.d. below the national mean because the MBE is her relative weakness). However, our candidate is good at essays and the MPT so her written score is 1 s.d. above the mean for her jurisdiction. According to the methodology that NCBE uses in scaling MBE scores, our candidate's essay score will be computed to be $140 + 15 = 155$ because the jurisdiction's MBE mean is 140 and its s.d. is 15. That would give our candidate a total UBE score of $155 + 125 = 280$, which is

¹⁹ *Id.* at 46, Table 1. The table entitled, Sample Essay Data Shown for Each Examinee, contains data for fifteen examinees.

high enough for admittance in several jurisdictions, including New Mexico, Idaho, Washington and New York.²⁰

Let's consider what happens if the jurisdiction's MBE mean is down at 135, with a standard deviation still at 15. If our candidate "scores 1 s.d. above the mean on the written, then her written score will be standardized to $135 + 15 = 150$. That means that her total UBE score would be $150 + 125 = 275$. She would no longer be eligible in Idaho (where the minimum required is 280) simply because of the slightly lower mean but same variance in MBE scores in her jurisdiction. Her skill level did not change: that of the pool of candidates did. Is this what we want to mean when we tout the "portability" of the UBE?"²¹

Now consider that the jurisdiction's MBE mean is at 140²² but the standard deviation is not as large — make it 12 rather than 15. The MBE score is still 125 but now our candidate's written score that is 1 s.d. above the mean in her jurisdiction gets scaled to $140 + 12 = 152$. Her total score on the UBE is then $152 + 125 = 277$ and again she would not be able to transport that score to Idaho for admission.

But those are pretty simplistic examples. If our candidate is really that good at the written component (in the 84th percentile in her jurisdiction if she is 1 s.d. above the mean) and she chooses a jurisdiction where the applicant pool is, for whatever reason, weaker in written performance, then her performance will be more than 1 s.d. higher in that jurisdiction. It can get a bit complicated to estimate this but just say that the MBE mean is down at 135 as in the second example, and relative to the weaker pool her

²⁰ E-mail from Nancy E. Johnson to Suzanne Darrow-Kleinhaus, Professor of Law and Director of Academic Development and Bar Programs, Touro Law Center (Feb. 14, 2016, 11:26 a.m. EST) (on file with author).

²¹ *Id.*

²² Case, *Demystifying Scaling to the MBE*, *supra* note 12, at 46, Table 1.

written score winds up being 2.5 s.d. above the mean. Then her written score would scale to $135 + 22.5 = 157.5$ and that elevates her total UBE score to $125 + 157.5 = 282.5$. This would give her entry into just about any UBE jurisdiction.²³

It would seem likely that with smaller sample sizes, it would be more likely to see variations from the normal distribution. However, it is not possible to determine how seriously that would distort the standardization because so little information about the national sample and the individual jurisdictions are available. Nonetheless, it is possible to see that the more you “work the numbers” the way that NCBE does,²⁴ the more you see that the same skill level could result in different UBE scores, depending on where the candidate takes the exam and what that jurisdiction’s applicant pool does on that particular exam, in terms of both skill level and also the range or spread of their scores.

“UBE shopping” may make a UBE score “fair” for the examinee

“Forum shopping,” however, may level the playing field for the individual in a way that the current scoring and weighting of the bar exam components does not. While relative grading may make a UBE score “unreliable” as to the “receiving” jurisdiction, it may make the UBE “fair” to the individual. By having a choice among UBE jurisdictions for where to take the exam, an examinee who performs better on the written component can compensate for a weaker MBE score by having that written score ranked and scaled in a “weaker” jurisdiction.

²³ E-mail from Nancy E. Johnson to Suzanne Darrow-Kleinhaus, Professor of Law and Director of Academic Development and Bar Programs, Touro Law Center (Feb. 14, 2016, 11:26 a.m. EST) (on file with author).

²⁴ Case, *Demystifying Scaling to the MBE*, *supra* note 12, at 46, Table 1.

This option is necessary to provide a “fair” path to bar passage when the UBE’s equal weighting of the MBE and written portion provides only a “fair system overall.”²⁵ In explaining what it means to be fair “overall”, Dr. Case notes that “[w]hile research has not shown that any ethnic or racial group performs better as a group on one format or the other, individuals may perform relatively better on one of the formats (i.e., some individuals perform better on the multiple-choice component whereas others perform better on the written components.”²⁶ Accordingly, a “single total scaled score allows examinees who perform better on one component to compensate for weaker performance on another component, and weighting the written and multiple-choice portions equally assures overall fairness.”²⁷

However, what Dr. Case fails to mention is that weighting bar exam components equally does not assure “overall fairness” for the individual who performs better on the written component when that score is the result of relative grading. Perhaps having an option for where an examinee takes the UBE will remedy the unfairness by allowing the individual’s strength on one component to have a “fair” chance of compensating for a weaker performance on another.

On the other hand, how “fair” is it to the other examinees in the UBE jurisdiction to where the score is transported? While essay grading by rank-ordering is considered a “grading fundamental”²⁸ and practiced within non-UBE jurisdictions as well as UBE jurisdictions, it has different implications in a UBE setting. Even assuming that in both a non-UBE jurisdiction and a UBE jurisdiction, an examinee’s raw scores on the written

²⁵ Case, *What’s In It For Me?*, *supra* note 3, at 50.

²⁶ *Id.* at 50-51. NCBE’s claim that “research has not shown that any ethnic or racial group performs better as a group on one format or the other” is disputed.

²⁷ *Id.* at 50-51.

²⁸ Gundersen, *supra* note 15, at 54.

portion are added up and scaled to the MBE mean and standard deviation for that jurisdiction, the difference is that the non-UBE earned score remains in that jurisdiction: it is not transferred for admission to practice law in another jurisdiction where a completely different group of candidates sat for the bar exam. The examinee with the “portable score” was not “ranked” against these examinees to achieve her score.

NCBE claims that the UBE’s consistency will make the bar admission process more comprehensible to the public.

NCBE’s claim that the UBE’s consistency will make the bar admission process more understandable to the public is insupportable when much of that process remains hidden from public scrutiny. NCBE is not making the bar admission process more comprehensible to the public when it speaks in hypotheticals even as it purports to “unlock” the mysteries of scaling essay scores to the MBE.

While NCBE releases the national MBE mean following each administration of the bar exam and some jurisdictions release their individual MBE mean, there is a general absence of information regarding the mean and standard deviation for the MBE and the written component used to determine bar scores in jurisdictions. Without this information, there is no way to replicate and therefore validate the “equating process” followed by NCBE and jurisdictions in arriving at examinee scores. Nor is there any way to assess the “validity and reliability of using only multiple choice items as anchors to equate forms of a mixed-format test.”²⁹

²⁹ Posting of Nancy E. Johnson to asp-1@chicagokent.kentlaw.edu (Apr.17, 2015, 1:44:50 p.m. EST) (The subject heading of this email is [ASP-L:5369] Re: NCBE Responses RE July 2014 MBE Nationwide Decline) (on file with author).

Recent studies in this area indicate a cause for concern as to whether NCBE's equating method works equivalently for different subpopulations.³⁰ In a 2011 study, Kim and Walker “looked at linking mixed-format using a multiple-choice anchor and asked whether it would produce comparable results for men and women. They found that when the correlation between the multiple choice and the written (constructed response items) is relatively low, large differences are seen between groups, and the use of multiple choice anchors is of questionable efficacy.”³¹

Since this study, other relevant works have been published. One is an empirical study by the same authors, Kim and Walker, where their main finding about the use of only multiple-choice anchor items is “[i]n the MC-only condition, the difference between the subpopulation functions and the total group function was not trivial in a score region that included cut scores, leading to inconsistent pass/fail decisions for low-performing examinees in particular....The research reinforces subpopulation invariance indices as a means of determining the adequacy of the anchor.”³² Commenting on this finding, Dr. Johnson notes that it is stunning in its implications in that “as more of the pool approaches the region we would call low-performing, bias increases.”³³

Equally concerning is the validity and adequacy of using only multiple choice items as anchors to equate forms of a mixed-format test — one that consists of essays

³⁰ *Id.*

³¹ *Id.* See also, SOOYEON KIM & MICHAEL E. WALKER, ETS, RESEARCH REPORT ETS RR-11-44, DOES LINKING MIXED-FORMAT TESTS USING A MULTIPLE-CHOICE ANCHOR PRODUCE COMPARABLE RESULTS FOR MALE AND FEMALE SUBGROUPS?, (2011) <https://www.ets.org/Media/Research/pdf/RR-11-44.pdf>.

³² Sooyeon Kim and Michael E. Walker, *Determining the Anchor Composition for a Mixed-Format Test: Evaluation of Subpopulation Invariance of Linking Functions*, 25 APPLIED MEASUREMENT IN EDUCATION 178 (2012).

³³ Posting of Nancy E. Johnson to asp-1@chicagokent.kentlaw.edu (Apr.17, 2015, 1:44:50 p.m. EST) (The subject heading of this email is [ASP-L:5369] Re: NCBE Responses RE July 2014 MBE Nationwide Decline) (on file with author).

and performance tests as well as multiple choice questions, as does the UBE.³⁴

According to Dr. Johnson:

when the correlation between multiple choice and a different format item is relatively low, significant differences in accuracy of equating are seen between men and women, and the use of multiple choice items as anchors is of questionable efficacy (Kim & Walker, 2011), presumably because the two formats are not measuring the same underlying ability. Susan Case ... reported the correlation between MBE and MPT to be down at .38, which may be the cause for concern in many jurisdictions. Very large sample sizes do not cure the problem.³⁵

However, whenever NCBE is questioned about the “reliability, validity, integrity, and fairness of the test and the processes by which it is created and scored,”³⁶ it appears to have but one answer — trust us because we ran the tests and we say that they are reliable. When NCBE informs the public that its test instruments are valid and reliable, we have only its word for it because NCBE does not share how it verifies its own questions — just that it does.³⁷

And it’s not like NCBE has not been asked. In response to the legal academy’s questioning of the MBE in light of the decline in the mean score for the July 2014 administration of the Multistate Bar Examination, Erica Moeser, President of the

³⁴ N.E. Johnson, Comment to *No the MBE was not “harder” than usual*, EXCESS OF DEMOCRACY (Sept. 28, 2015), <http://excessofdemocracy.com/blog/2015/9/no-the-mbe-was-not-harder-than-usual#comments-552c15c9e4b06d12aaa571d0=>.

³⁵ *Id.*, See also Susan M. Case, *The Testing Column, Relationships Among Bar Examination Component Scores: Do They Measure Anything Different*, THE BAR EXAMINER, Aug. 2008, at 31, http://www.ncbex.org/pdfviewer/?file=%2Fassets%2Fmedia_files%2FBar-Examiner%2Farticles%2F2008%2F770308_testing.pdf. With respect to the correlation between MBE and MPT scores for the data set, Dr. Case wrote that “the correlation with the MBE is 0.55 for the local essay questions, 0.58 for the MEE, and 0.38 for the MPT. This shows a moderate correlation for both the locally developed essay questions and the MEE, but a weaker correlation for the MPT, indicating that the MPT is measuring different skills than the MBE, and that the MPT skills are less like those measured by the MBE than are the skills measured by the MEE and local essay questions.” If the correlation between MBE and MPT scores is so low, then how can scaling the MPT to the MBE be a reliable measure of anything, let alone an examinee’s skills?

³⁶ Erica M. Moeser, *President’s Page*, THE BAR EXAMINER, Mar. 2015 at 4, http://www.ncbex.org/assets/media_files/Bar-Examiner/articles/2015/840115-abridged.pdf.

³⁷ Erica M. Moeser, *President’s Page*, THE BAR EXAMINER, Dec. 2015 at 4, http://www.ncbex.org/pdfviewer/?file=%2Fassets%2Fmedia_files%2FBar-Examiner%2Fissues%2F2015-December%2FBE-Dec2015-PresidentPage.pdf.

National Conference of Bar Examiners, wrote that “we are confident of the correctness of the scores as reported. Because of the importance of getting things right, we engaged in more replications of our equating procedure internally — and indeed, more review of our procedures for selecting test items — than usual. Had we detected error, we would have reported and acted upon it. We found no error.”³⁸

Moeser further explained that “in an excess of caution we then took the additional step of engaging the Center for Advanced Studies in Measurement and Assessment (CASMA) of the University of Iowa to review our results.”³⁹ She reported that CASMA replicated NCBE's results “and confirmed the results using alternative means of conducting the equating.”⁴⁰ Noting that the report was “confidential,” Moeser quoted from it that “[t]he results for [the replication] were almost identical to those provided by NCBE” and that it “did not find any explanation for the mean drop observed in July 2014 attributable to the statistical equating procedures examined in this report.”⁴¹ As Moeser concluded, “case closed.”⁴²

Scoring issues with UBE’s “portability” may make admissions committees vulnerable to legal claims.

There is a difference between an exam score earned by an examinee in an individual jurisdiction scoring its own written exam and an exam score earned in one jurisdiction that is “transported” to another. Even if the UBE is uniformly administered,

³⁸ Erica M. Moeser, *President’s Page*, THE BAR EXAMINER, June 2015, at 4, http://www.ncbex.org/pdfviewer/?file=%2Fassets%2Fmedia_files%2FBar-Examiner%2Farticles%2F2015%2F840215-PresidentPage.pdf.

³⁹ *Id.*

⁴⁰ *Id.*

⁴¹ *Id.* at 4-5.

⁴² *Id.* at 5.

graded, and scored by the jurisdictions that adopt it, we have seen how it is possible that a 280 score in one jurisdiction is not the same as a 280 score in another.

Let's consider our hypothetical candidate once again. Suppose she takes the bar exam in one UBE jurisdiction and scores a 278. She understands how close she is to the magic 280 that would allow her admittance to Idaho, and she understands that the scaling seems weird. Could she not file a discrimination suit, based on NCBE's scoring practices⁴³ and seek discovery to force release of information about the mean and standard deviation for the MBE and the written score in each of the two jurisdictions? How long will it be before a disappointed examinee challenges the portability of a UBE score?

Now that we know the UBE can result in a different numerical score for the exact same performance depending on where the examinee wrote the test, what we decide to do next is critical. Of course we can ignore what we know and allow bar candidates to "UBE shop." Or we can insist that a "uniform score" be truly uniform. The "only way for the UBE to be truly portable is to get every jurisdiction to agree to use and pay a centralized scoring service to grade it and standardize it based wholly on a national distribution. That scoring service would, of course, be NCBE."⁴⁴

Before we proceed down that road, however, we need to ask the following questions. They are important but the answers are even more important because they determine the future of legal education and access to the profession.

- Over what aspects of the licensing process do we want "centralized control?"

⁴³ Case, *Demystifying Scaling to the MBE*, *supra* note 12.

⁴⁴ E-mail from Nancy E. Johnson to Suzanne Darrow-Kleinhaus, Professor of Law and Director of Academic Development and Bar Programs, Touro Law Center (Feb. 14, 2016, 3:24 p.m. EST) (on file with author).

- Do we want a “central collection point” for all bar exam data for all bar candidates as NCBE has offered to become?⁴⁵ Is NCBE the right entity for this purpose? If so, what oversight shall there be and by whom?
- Is the next step a national law license?

Finally, the very question of whether the UBE achieves its primary purpose of assessing whether a candidate is competent to practice law is in doubt. As presently conceived and administered, the UBE cannot be a measure of a candidate’s “minimum competency” if the same person can be found “competent” to practice law in one UBE jurisdiction and “incompetent” in another when it is the same person with the same skill level writing the same exam.

⁴⁵ Erica Moeser, *President's Page*, THE BAR EXAMINER, Nov. 2009 at 5, www.ncbex.org/pdfviewer/?file=%2Fassets%2Fmedia_files%2FBar-Examiner%2Farticles%2F2009%2F780409_PresidentsPage.pdf.