IN THE SUPREME COURT OF IOWA

The Iowa Board of Law Examiners' Recommendation to the Supreme Court of Iowa to Amend Iowa Court Rule 31.11 to Remove the Automatic Review Process for the Written Components of the Iowa Bar Examination

June 2018

The Iowa Board of Law Examiners:

Lora L. McCollom, Chair

A. Eric Neu, Vice Chair

Jean Z. Dickson

Stacey N. Warren

Timothy J. Luce

Patricia Thayer

R. Keith Richardson

I.   Introduction

Prior to 2012, Iowa Court Rule 31.11 provided that individuals who failed the Iowa bar examination a single time had no right to appeal their score, but that an individual who had failed two or more exams and had scored 260 points or higher on the exam at issue had the right to appeal as to the Board of Law Examiners' scoring of the written portion of his or her exam.[1]   The applicant was given an opportunity to choose no more than four of his or her written answers for review.[2] In February of 2012, the Supreme Court entered an order amending Rule 31.11 to do away with the old process and to implement a new appeal process which remains largely in place. The new appeal process would be an automatic one—any applicant whose combined, scaled score on the bar exam was between 260 and 265 was given an automatic review of the written portion of his or her exam.[3] Those appeal procedures will be outlined in more detail below.

Also in 2012, a provision for Supreme Court review of the score ultimately settled on by the board was put into place where an applicant could petition the court for review his or her answers if the score remained in the 260- to 265-range.[4] Currently, Rule 31.11(3) does provide a right for applicants whose score is between 260 and 265 to petition to the Supreme Court requesting review of the board's score determination. However, the board's decision regarding a score

---

[1] *See* Iowa R. Ct. 31.11 (2011).
[2] *Id.*
[3] *See* Iowa R. Ct. 31.11(1)–(2) (2012).
[4] *See* Iowa R. Ct. 31.11(3) (2012).

may only be reviewed under "extraordinary circumstances," which include "issues such as the board's refusal to correct a clear mathematical error, but would not include a claim that the board erred in the grade assigned to a particular answer."[5]

On August 4, 2015, the Iowa Supreme Court adopted the Uniform Bar Examination (UBE) as the admissions examination for this state. The first administration of the UBE was February 2016. At the same time the UBE was adopted, rule 31.11(1) was amended to clarify that the board will not conduct any review of written scores after the bar examination results are released. That provision remains in effect, along with the automatic review of the written responses outlined above.

The National Conference of Bar Examiners (NCBE) is the entity that creates the written and multiple choice items found on the bar exam. All UBE jurisdictions have agreed to "train and calibrate their MEE and MPT graders to follow uniform standards applied by all UBE jurisdictions."[6] There are certain policies followed by each jurisdiction that has adopted the UBE "in order to produce comparable scores, enhance score portability, and ensure reliable transfer of scores."[7] Although UBE jurisdictions are allowed to regrade the written components of the bar exam after the NCBE has performed its initial

---

[5] Iowa R. Ct. 31.11(3) (2018). The board is not recommending removal of this provision related to supreme court review contained within rule 31.11(3).

[6] *See* Exhibit A, National Conference of Bar Examiners, *Understanding the Uniform Bar Examination*, updated Oct. 12, 2017.

[7] Kellie R. Early, *The UBE: The Policies Behind the Portability*, THE BAR EXAMINER, September 2011 at 17, 17.

scaling and combining of written and Multistate Bar Exam (MBE) scores,[8] the NCBE does not recommend this practice. UBE jurisdictions, including Iowa, have agreed not to regrade written answers after examination results have been posted.[9] The NCBE has developed recommendations for jurisdictions to follow in the event they allow for regrading of the written components of the bar exam in order ensure consistency with the original grading standards is maintained.[10]

There have been five administrations of the UBE in Iowa since its implementation. At its score-setting meeting in conjunction with the February 2018 bar exam, the board discussed the automatic appeal process and voted to recommend to the Supreme Court that the automatic review procedures set forth in Rule 31.11(1)–(2) be removed and that the appeal procedure contained in rule 31.11(3), related to Supreme Court review, be retained.

II.   Scoring the UBE

The UBE can be divided into two components: the MBE and the written components, which are comprised of the Multistate Essay Examination (MEE) and the Multistate Performance Test (MPT). The MBE is a 200-question multiple choice examination generally taken on the Wednesday of the exam administration. Scores can range from 40 to 200 and are calculated by the NCBE

---

[8] The Multistate Bar Exam consists of 200 multiple-choice questions and is administered as part of the Uniform Bar Exam

[9] Early, *supra* note 7, at 19.

[10] *See id.*

based on a statistical process known as equating.[11] In UBE jurisdictions, the MBE accounts for 50% of the applicant's score.[12]

The MEE consists of six 30-minute essay questions, and is generally given on the Tuesday afternoon of the exam administration.[13] The MEE accounts for 30% of the applicant's score in UBE jurisdictions, including Iowa.[14] The MPT consists of two 90-minute skills tests.[15] It represents 20% of the applicant's total UBE score.[16]

The scoring process of the written items is as follows: graders assign each of the six MEE items and two MPT items written answer a score. These written scores are combined to create a raw score for the applicant. The raw scores are then sent to the NCBE in order for the total, raw written score to be converted to a scaled score. The highest total raw written score is assigned the same score as the highest MBE score, the second highest raw score is assigned the same score as the second highest MBE score, and this continues until the lowest raw written score is assigned the same score as the lowest MBE score. In that sense, the written portion of the examination is scaled according to the MBE.

All UBE jurisdictions train and calibrate their MEE and MPT graders to follow uniform standards established by the NCBE.[17] "Calibration is the process

---

[11] http://www.ncbex.org/exams/mbe/scores/
[12] *Id.*
[13] http://www.ncbex.org/exams/mee/
[14] *Id.*
[15] http://www.ncbex.org/exams/mpt/
[16] *Id.*
[17] *See* Exhibit A.

of developing coherent judgment in assigning points, using the standards set out in the grading materials, so that the rank-ordering of answers is done consistently over the entire course of grading."[18] Instead of assigning a letter grade or a "passing" or "failing" score to the applicants' MEE and MPT responses, graders are trained to evaluate examinees' answers on a relative basis.[19] Now-President of the NCBE, Judith Gundersen explained relative grading and the rank-ordering process:

> Relative grading training helps graders identify consistent standards in ranking papers and then apply those standards to put papers in piles according to their relative strength. The 1–6 scale used at the [NCBE grading] workshop simply means that a score of 6 is reserved for the best papers among all answers assigned to a particular grader. It is better than a 5, which is better than a 4, and so on, all the way to 1—a paper that is among the weakest papers. Relative grading means that in any group of answers, even if no single paper addresses all the points raised in an item, the strongest papers still deserve a 6 (using a 1–6 score scale). They do not have to be perfect nor necessarily deserve a true A or 100% . . . . Using the same principles, a paper need not be completely devoid of content to get a 1 if the other papers are strong.[20]

The use of relative grading is backed by various psychometric and policy reasons.[21]

III.   Scoring the MEE and MPT in Iowa

As a jurisdiction that has adopted the UBE, Iowa adheres to the scoring procedures outlined above. Attorney board members and team leaders[22] are

---

[18] Early, *supra* note 7, at 18.
[19] *See* Judith A. Gundersen, *It's All Relative—MEE and MPT Grading, That Is*, THE BAR EXAMINER, June 2016 at 37, 38.
[20] *Id.*
[21] *Id.* at 39–43.
[22] These Team Leaders are temporary examiners appointed by the supreme court pursuant to Rule 31.1(1)(*d*).

placed in charge of the grading of one MEE or one MPT item. They participate in a calibration session with the NCBE, either in person or remotely. Each of these board members or team leaders is placed in charge of a team of Iowa attorneys who are assigned to an MEE or MPT item and appointed by the Supreme Court as temporary examiners.[23] The team members work in groups of two and are tasked with assigning a score to each item according to the rank-ordering scoring system mentioned above, and the board member/team leader works with team members to ensure calibration throughout the grading session. By using this method, each written answer is reviewed by two Iowa attorneys and assigned a score based upon the calibration led by the board member/team leader. This procedure is aligned with the recommended practice of the NCBE Testing and Research Department.[24] Attorney board members, team leaders, and graders all gather in person the month following the administration of the bar examination for a two-day grading session. The first day of the session consists of an orientation as well as individual board members and team leaders conducting calibration sessions with their individual teams. Teams of two examiners then spend the second day reading applicants' answers to MEE or MPT items, depending on the team they are assigned to, and assigning scores consistent with the calibration session and the procedures outlined above. Conducting the grading in this manner comports with the NCBE's standard of graders being calibrated throughout the grading process. It also is in line with

---

[23] *See* Rule 31.1(1)(*d*).

[24] Mark A. Albanese, Ph.D., *The Testing Column: Regrading Essays and MPTs—And Other Things That Go Bump in the Night*, THE BAR EXAMINER, March 2016 at 58, 61.

the recommended practice by NCBE Testing and Research Department as a means of avoiding misgrading, because grading is being conducted "in a time period that is as consolidated as possible."[25]

In Iowa, the grading scale for MEE and MPT components is one to six, with a six representing the highest score possible on that individual essay or performance item. Instead of a system where a given number represents a "passing" score for an essay or performance item, the score assigned to each individual item represents where the individual's answer falls within the scope of all of the answers submitted in Iowa. For example, answers receiving a six represent roughly the top one-sixth of the answers for that particular item, while the answers that receive a one represent roughly the bottom one-sixth of the answers for the item.

In short, each written item is rank-ordered on a scale of one to six, and the applicant's written scores for each of the six MEE items and the two MPT items are added together to create a raw score for each applicant. Those total raw scores are then rank-ordered again and scaled against the scaled MBE scores.

IV.   The Current Automatic Review Process in Iowa

Applicants who obtain scaled scores between a 260 and 265 are entitled to automatic reviews of their written answers in the MEE and MPT components

---

[25] *Id.* at 58.

of the bar examination prior to the exam results being released.[26] Rule 31.11(2)

sets forth the procedures for automatic review as follows:

> a. The attorney members of the board and any temporary examiners the board may designate will review the applicant's written answers. The answers will be submitted on an anonymous basis without oral argument or hearing. If it appears that an answer should receive a different score (whether higher or lower), that score will be used to determine the applicant's scaled score. The board will maintain a record of any changes made to the scoring of the individual questions on review.
>
> b. Following its review, the board will recommend to the supreme court that the applicant be admitted to the practice of law in Iowa if the applicant's combined, scaled score after review is at least 266. An applicant whose combined, scaled score after review is 265 or below will be deemed to have failed the examination.[27]

When an applicant's test qualifies for automatic review, the board

member/team leader is the individual who conducts the review. The board

member/team leader reviews the answers corresponding to his or her assigned

MEE or MPT question for each of the applicants whose exam qualifies for

automatic review. These answers are submitted to the board member/team

leader anonymously. The board member/team leader is also given benchmark

responses that fit into each scoring level for his or her MEE or MPT component.

This results in each board member/team leader receiving six benchmark

papers—one at each level on the scale of 1–6. These benchmarks are selected

by the grading teams as being illustrative of answers that fall within that score

---

[26] Iowa Ct. R. 31.11(1).
[27] Iowa Ct. R. 31.11(2).

level, and the benchmark answers are never from applicants who fall within the automatic review process.

Every MEE and MPT component is reviewed, and a new raw score is generated if there are any changes made by the board member/team leader. These new raw scores are resubmitted to the NCBE for rescaling in conjunction with the procedures outlined above. The current review procedure can result in an additional one to two weeks being added to the grading process.

V.   NCBE's Position

While regrading is permitted, the NCBE does not recommend it.[28] Kelly Early, the Director of Administration for the NCBE detailed several flaws in regrading processes in relation to post-score-release regrading that can also be applied to an evaluation of pre-score-release regrading:

> Regrading is not likely to produce psychometrically sound scores if the pass/fail status of the applicants is known, if the original scores are known, if the regrading is done remote in time from the original grading of the entire pool of answers, I only failing answers are reviewed, and/or if scores are only increased and never decreased. Calibration or consistency with the grading standards is difficult to maintain under such circumstances.[29]

In a March 2016 article of *The Bar Examiner*, Director of Testing and Research for the National Conference of Bar Examiners Mark Albanese outlined various types of misgrading that can occur in conjunction with the bar

---

[28] *See* Albanese, *supra* note 24, at 61.
[29] Early, *supra* note 7, at 19.

examination and mentioned multiple solutions.[30] The best remedy, he noted, is

to "avoid misgrading in the first place" by utilizing "well-trained and calibrated

graders who complete their evaluations in a time period that is as consolidated

as possible."[31] He emphasized the importance of grader training, both from the

NCBE as well as the individual jurisdictions. He categorized regrading only the

papers of those examinees who failed to achieve a passing score as the "least

best" remedy to misgrading[32] and outlined the inherent bias an individual

regrading these papers.[33] After setting out recommendations on best practice in

the event a jurisdiction does regrade MPT and MEE answers, Albanese

summarized NCBE's recommendation:

> NCBE does not recommend regrading any part of the bar examination. We strongly recommend that jurisdictions put their entire grading resources into making certain that the initial grades awarded are of the highest possible caliber. We put our money where our mouth is and invest substantial resources and effort into providing grading workshops after each bar examination. Best practice, and what is probably the industry standard, is to have two graders grade each essay.[34]

VI.   Other Jurisdictions' Practices

Other jurisdictions are split on whether they regrade the written

components of the bar exam. In recent surveys, eleven of the responding UBE

jurisdictions, including Iowa, indicated it was their practice to engage in some

---

[30] Albanese, *supra* note 24, at 58.

[31] *Id.*

[32] *Id.* at 58–59.

[33] *Id.* at 60 ("If the policy is to have all examinees who fail regraded, then graders will know going into the regrading process that they are looking at only the essays of examinees who failed in the first round of grading. Depending upon a grader's perspective, whether sympathetic or antipathetic to the failing examinee's plight, his or her grading may be biased.").

[34] *Id.* at 61.

sort of pre-score-release-regrading. Eleven of the responding UBE jurisdictions indicated they did not regrade. In the Midwest, Kansas, Nebraska, and North Dakota were the responding jurisdictions that do not regrade, while Minnesota and Missouri indicated they do so.

VII.   Conclusion and Recommendation

After considering its experience with the Uniform Bar Examination in this state since its implementation, including the grading procedures for the written components of the exam and current regrading system, the board recommends the court adopt NCBE's recommendation to eliminate the automatic regrading process in this state. The board is confident the extensive resources, time, and effort that go into the initial grading process, including utilization of NCBE's grading workshop for calibration of board members and team leaders, an intensive calibration among Iowa MEE and MPT grading team members, continued involvement of board members and team leaders in the calibration process, and a two-day grading process where two grading team members review and score each MEE and MPT answer following uniform standards applied by all UBE jurisdictions, ensure a fair and accurate system of scoring the written components of the UBE. The board believes that the rank-ordering scoring system and necessary grader calibration are best preserved with the ranking and scoring system performed during the primary grading session, with all answers being considered at one time rather than during a later regrading session where only a limited number of answers are considered. Finally, elimination of the

automatic appeal process while ensuring fair, consistent, and accurate grading of the written components of the bar exam will shorten the time between administration of the bar exam and release of the results. In turn, this will help facilitate admissions ceremonies being conducted sooner in relation to the administration of the bar exam and benefit examinees in allowing them to begin their careers as Iowa-licensed attorneys sooner.

EXHIBIT "A"

National Conference of Bar Examiners, *Understanding the Uniform Bar Examination* (Oct. 12, 2017)

THE UBE
Multistate Uniform Bar Examination
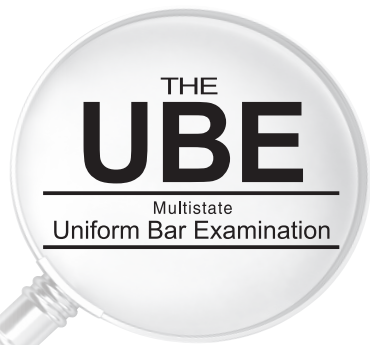
Understanding the Uniform Bar Examination

What Is the UBE?

The Uniform Bar Examination (UBE) is prepared and coordinated by the National Conference of Bar Examiners to test knowledge and skills that every lawyer should be able to demonstrate prior to becoming licensed to practice law. It is composed of the Multistate Essay Examination (MEE), two Multistate Performance Test (MPT) tasks, and the Multistate Bar Examination (MBE). It is uniformly administered, graded, and scored by user jurisdictions and results in a portable score.

The UBE is administered over two days, with the MBE given on the last Wednesday of February and July and the MEE and MPT given on the Tuesday prior to that. The MEE and MPT scores are scaled to the MBE, with the MBE weighted 50%, the MEE 30%, and the MPT 20%.

Jurisdictions that use the UBE continue to

- decide who may sit for the bar exam and who will be admitted to practice.
- determine underlying educational requirements.
- make all character and fitness decisions.
- set their own policies regarding the number of times candidates may retake the bar examination.
- make ADA decisions.
- grade the MEE and MPT.
- set their own pre-release regrading policies.
- assess candidate knowledge of jurisdiction-specific content through a separate test, course, or some combination of the two if the jurisdiction chooses.
- accept MBE scores earned in a previous examination or concurrently in another jurisdiction for purposes of making local admission decisions if they wish. Note: candidates must sit for all portions of the UBE in the same UBE jurisdiction and in the same administration to earn a portable UBE score.
- set their own passing scores.
- determine how long incoming UBE scores will be accepted.
- maintain the security of test content and provide appropriate testing conditions by administering the UBE at specified times and in accordance with the rules laid out in the NCBE Supervisor's Manual, including the guidelines for room setup, book distribution, seating charts, and proctor selection and training.

To ensure that candidates are assessed consistently across jurisdictions, UBE jurisdictions will

- administer the entire examination (MEE, MPT, and MBE) to each UBE candidate in the same administration. Banked, transferred, or concurrent MBE or written scaled scores earned in a prior examination or concurrently in another jurisdiction may not be used in calculating UBE total scores.
- grade the MEE and MPT using generally applicable rules of law rather than jurisdiction-specific law.
- train and calibrate their MEE and MPT graders to follow uniform standards applied by all UBE jurisdictions.
- have NCBE perform the scaling of the MEE and MPT scores to the MBE to ensure that score calculations are performed consistently across jurisdictions.
- agree that, to qualify as UBE scores, the scores will not be changed after examination results have been announced.
- report on their test administrations and permit occasional audit by NCBE to verify that best practices are being followed.

To facilitate score portability and transfers, UBE jurisdictions will

- generate a UBE total score expressed as a whole number on a 400-point scale.
- accept transferred UBE scores that meet their passing standard whether or not the score meets the passing standard in the testing jurisdiction.
- require candidates to provide sufficient information on the MBE answer sheets to identify their scores for transfer by NCBE, including the candidate's name, date of birth, NCBE Number, and the last four digits of the Social Security number.
- submit all UBE scores to a central registry maintained by NCBE to ensure that a full score history is reported by NCBE to receiving jurisdictions when candidates request UBE score transfers.
- provide, or have NCBE provide, candidates with their written scaled scores, MBE scaled scores, and UBE total scores so that candidates can determine if their scores are high enough to transfer to other jurisdictions.

Role of the Jurisdictions

Since the UBE's inception, jurisdictions have been actively involved in shaping UBE policies and implementing best practices. Representatives from UBE jurisdictions regularly discuss administrative issues and make recommendations to NCBE's Board of Trustees with respect to policy decisions related to the UBE. A representative from each UBE jurisdiction serves on NCBE's Special Committee on the Uniform Bar Examination.

2017–2018 NCBE Special Committee on the Uniform Bar Examination

Hon. Cynthia L. Martin, Chair
Judge, Missouri Court of
    Appeals for the Western
    District
Kansas City, MO

Hon. Barry Anderson
Associate Justice, Minnesota
    Supreme Court
St. Paul, MN

Jennifer Archer
Chair, Maine Board of
    Bar Examiners
Portland, ME

Hon. Geoffrey W. Barnard
U.S. Magistrate Judge,
    Supreme Court of the Virgin
    Islands
St. Thomas, VI

Hon. Donald W. Beatty
Chief Justice, South Carolina
    Supreme Court
Columbia, SC

Gary W. Bjelland
Vice-Chair, Montana Board
    of Bar Examiners
Great Falls, MT

Maureen Ryan Braley
Director of Admissions,
    Idaho State Bar
Boise, ID

Hon. Mark S. Cady
Chief Justice, Iowa Supreme
    Court
Des Moines, IA

Hon. Edward L. Chavez
Justice, New Mexico Supreme
    Court
Santa Fe, NM

Hon. Linda Stewart Dalianis
Chief Justice, Supreme Court
    of New Hampshire
Concord, NH

Anne C. Dranginis
Chair, Connecticut Bar
    Examining Committee
Bridgeport, CT

Hon. Christine M. Durham
Justice, Supreme Court of Utah
Salt Lake City, UT

Hon. Zel M. Fischer
Chief Justice, Supreme Court
    of Missouri
Jefferson City, MO

Hon. Michael G. Heavican
Chief Justice, Supreme Court
    of Nebraska
Lincoln, NE

Christina Holland
Vice Chairman, Kansas Board
    of Law Examiners
Kansas City, MO

Daniel F. Johnson
Chair, Alabama Board of
    Bar Examiners
Dothan, AL

Stephen Douglas Juge
Member, District of Columbia
    Committee on Admissions
Washington, DC

Keith J. Kasper
Chair, Vermont Board of
    Bar Examiners
Burlington, VT

Laura Maresca
Vice-Chair of the Colorado
    Board of Law Examiners,
    Law Committee
Denver, CO

Carol Mitchell
Attorney Admissions Manager,
    Attorney Admissions
    Supreme Court of Arizona
Phoenix, AZ

Hon. Susan Owens
Justice, Washington Supreme
    Court
Olympia, WA

Nicole L. Robbins
Vice-Chair, Oregon Board
    of Bar Examiners
West Linn, OR

Darin B. Scheer
Chair, Wyoming Board of
    Law Examiners
Casper, WY

Carol P. Smith
Member, West Virginia Board
    of Law Examiners
Charleston, WV

Hon. Gerald W. VandeWalle
Chief Justice, North Dakota
    Supreme Court
Bismarck, ND

Elizabeth A. Weiler
Chair, New Jersey Board of
    Bar Examiners
Cranford, NJ

Marilyn J. Wellington
Executive Director,
    Massachusetts Board of
    Bar Examiners
Boston, MA

Bryan R. Williams
Member, New York State
    Board of Law Examiners
New York, NY

Hon. Daniel E. Winfree
Justice, Alaska Supreme Court
Fairbanks, AK

Inquiries regarding the UBE may be directed to Kellie Early at kearly@ncbex.org or 608-280-8550.

EXHIBIT "B"

Kellie R. Early, *The UBE: The Policies Behind the Portability*, THE BAR EXAMINER, September 2011

The UBE:
The Policies behind the Portability

by Kellie R. Early

T he Uniform Bar Examination (UBE) has moved from concept to reality with its adoption by Missouri, North Dakota, Alabama, Idaho, and Washington.[1] The UBE is made up of a common set of six Multistate Essay Examination (MEE) questions, two Multistate Performance Test (MPT) tasks, and the Multistate Bar Examination (MBE), and the exam results in a portable score.[2]

The UBE is more than just a shared set of test components. At its essence, it is an agreement to give full faith and credit to examination scores generated in participating jurisdictions based upon the fact that all UBE jurisdictions uniformly administer, grade, and score the same examination.

Certain policies are followed by UBE jurisdictions in order to produce comparable scores, enhance score portability, and ensure reliable transfer of scores. Jurisdictions agree to adhere to these policies in order to be recognized as UBE jurisdictions and generate scores that qualify to be certified by the National Conference of Bar Examiners (NCBE) as UBE scores. These policies define what the UBE is and, by extension, what it is not.

The UBE is not reciprocal admission. The only element of reciprocity in the UBE is score portability; that is, UBE jurisdictions must accept scores from other UBE jurisdictions. But it is only the score that is portable, not the applicant's status in the testing jurisdiction. The fact that an applicant passes the UBE in one jurisdiction and is admitted to practice there does not, alone, qualify the applicant for admission in other UBE jurisdictions. It remains the responsibility of each UBE jurisdiction to set the passing score that it concludes represents proof of minimum competence to practice law within its borders and to determine all other admission requirements. Jurisdictions that adopt the UBE are merely using the same high-quality examination to determine whether applicants have demonstrated the fundamental knowledge and skills necessary to begin practice. And because it is the same exam, it doesn't matter where that score was earned. This article discusses the policies that make the UBE work.

The Same Exam, Administered Consistently: The Policies

Standardized testing conditions contribute to score comparability. To ensure that testing conditions are as uniform as possible, UBE jurisdictions follow the instructions set out in the *Supervisor's Manual*[3] for administering the examination. The *Supervisor's Manual* prescribes procedures for, among other things, maintaining the security of testing materials, providing a suitable testing environment, deterring cheating, proctoring the examination, dealing with disturbances, and reporting any irregularities that occur in the administration of the exam.

In addition to the procedures provided in the *Supervisor's Manual*, which apply to all jurisdictions that use NCBE's tests, there are a couple of other procedures specific to administration of the UBE. First, UBE jurisdictions use a common set of six MEE questions, which are answered according to generally applicable principles of law rather than jurisdiction-specific law.[4] Second, UBE jurisdictions administer the two MPT tasks in one seamless three-hour test session rather than two 90-minute sessions.[5]

To earn UBE scores, applicants must sit for all portions of the examination in the same administration and cannot rely upon banked or transferred written-component or MBE scores from previous examinations taken in the testing jurisdiction or in other jurisdictions. Use of banked or transferred scores from prior examinations allows applicants to sit for only one day of the current examination, which is not as demanding as having to prepare for and take all components in a single administration.[6] In order for scores to be comparable, applicants must sit for all components in the same administration to earn a UBE score. UBE jurisdictions may continue to allow applicants to use banked or transferred scores to gain admission locally, but such applicants do not earn portable scores.

UBE jurisdictions continue to make their own decisions about whether to grant testing accommodations under the ADA, and NCBE plays no role in such decisions. NCBE has sponsored development of a model form that all jurisdictions (UBE and non-UBE) may opt to have applicants use to request test accommodations. Use of the model form should lead to greater consistency in the information and documentation supplied by applicants and considered by jurisdictions in making ADA decisions.[7]

The Same Exam, Graded Consistently: The Policies

The answers of applicants in each jurisdiction are graded within that jurisdiction using the general principles of law set out in the MEE and MPT grading materials prepared by NCBE. UBE graders must adhere to the grading rubrics set out in the grading materials so that the same weight is assigned by all UBE jurisdictions to the various issues tested by each question. UBE jurisdictions may continue to use whatever raw scale they wish in grading the MEE and MPT, because the raw scores are converted to the MBE scale.

> While calibration within each jurisdiction remains critical, it is not necessary to calibrate graders *across* UBE jurisdictions because the MEE and MPT scores are scaled to the MBE scores within each jurisdiction.

Further, UBE jurisdictions continue to calibrate their graders within the jurisdiction. Calibration is the process of developing coherent judgment in assigning points, using the standards set out in the grading materials, so that the rank-ordering of answers is done consistently over the entire course of grading, either by a single grader or, if more than one grader per question is used, by the multiple graders. While calibration within each jurisdiction remains critical, it is not necessary to calibrate graders *across* UBE jurisdictions because the MEE and MPT scores are scaled to the MBE scores within each jurisdiction.

NCBE provides educational opportunities for graders from all jurisdictions that use its tests. UBE jurisdictions, particularly those that have not previously used the MEE and/or MPT, are encouraged

to have their graders attend NCBE's grading workshop, either in person or by teleconference, the weekend following the examination.[8] Additionally, NCBE offers several other educational events that cover the topic of grading, among other relevant topics, on an annual or biannual basis, attendance at which is funded by NCBE for one or more representatives of UBE and non-UBE jurisdictions alike. NCBE is also available to consult directly with jurisdictions that request additional assistance with training graders.

Because one of the purposes of both the MEE and MPT is to test the applicant's ability to communicate effectively in writing, UBE jurisdictions take communication skills into consideration when grading the MEE and MPT. Applicants are expected to present a clear, concise, and well-organized composition and are expected to write in complete sentences, using appropriate grammar and syntax.[9] At this time, there is not a separate communication score or set percentage of points associated with communication skills; rather, communication skills are one aspect of the scores assigned to MEE and MPT answers.

UBE jurisdictions do not regrade the answers of failing applicants after examination results have been released.[10] Regrading is not likely to produce psychometrically sound scores if the pass/fail status of the applicants is known, if the original scores are known, if the regrading is done remote in time from the original grading of the entire pool of answers, if only failing answers are reviewed, and/or if scores are only increased and never decreased. Calibration or consistency with the grading standards is difficult to maintain under such circumstances. Because most, if not all, of these circumstances are present when regrading takes place after release of results, jurisdictions have good reason not to accept scores that are the result of post-release regrading.[11] Therefore,

UBE jurisdictions agree not to engage in post-release regrading. UBE jurisdictions may engage in pre-release regrading of answers, assuming that it is appropriately conducted to maintain consistency with the original grading standards, but once results are released, no further review of answers is undertaken.

THE SAME EXAM, SCORED CONSISTENTLY: THE POLICIES

NCBE performs scaling and combining of scores for all UBE jurisdictions to ensure consistency in how scores are calculated.[12] UBE jurisdictions provide NCBE with the raw scores for each of the six MEEs and two MPTs so that NCBE can make sure that the proper weighting is applied and that no scores from a jurisdiction-specific exam component are intermingled with UBE scores.[13]

The MEE is weighted 30%, the MPT 20%, and the MBE 50% in calculating the UBE total score. The written-component scores (MEE and MPT) are scaled to the MBE using the standard deviation method.

Uniformity in rounding of UBE scores is necessary for score comparability. The written-component and MBE scaled scores are rounded to one decimal; these two decimal scores are combined, and the UBE total score is rounded to a whole number and stated on a 400-point scale.[14]

ENSURING RELIABLE TRANSFER OF SCORES: THE POLICIES

NCBE serves as the central repository of UBE scores and performs score transfer services for all UBE jurisdictions.[15] When an applicant requests to transfer a UBE score, NCBE sends the receiving jurisdiction an official transcript of the applicant's full UBE score history across all jurisdictions and exam dates, with

the scores certified by NCBE. Those UBE jurisdictions that place a limit on the number of times an applicant may retake the examination will find this particularly important. In the context of the UBE, such jurisdictions may want to count all attempts to earn a score that is passing in the receiving jurisdiction, regardless of where the applicant tested, because the applicant took the same exam. For example, if Jurisdiction A's passing score is 266 and it limits attempts to three, and an applicant sits for the UBE four times in other jurisdictions, earning scores of 257, 259, 262, and 266, in that order, Jurisdiction A might refuse to accept the applicant's score of 266 because it was earned in the fourth attempt.[16] In order that complete score histories can be provided in the score transcripts, all UBE jurisdictions agree that NCBE is the central repository and sole transferor of certified UBE scores.

To create accurate transcripts for applicants who take the UBE multiple times or in multiple jurisdictions, NCBE must have sufficient biographical data[17] to tie all the scores together. Thus, UBE jurisdictions agree to instruct applicants to provide the necessary identifying information on their MBE answer sheets.[18]

ENHANCING SCORE PORTABILITY: THE POLICIES

UBE jurisdictions provide, or allow NCBE to provide,[19] each applicant with his UBE scores (MEE/MPT scaled score, MBE scaled score, and UBE total score) so that applicants can determine whether they meet the minimum passing score requirements of other jurisdictions. Although some non-UBE jurisdictions restrict reporting of scores to applicants,[20]

> . . . [T]HE UBE IS THE SAME EXAM, ADMINISTERED, GRADED, AND SCORED UNIFORMLY BY ALL UBE JURISDICTIONS. THEREFORE, THE SAME REQUIREMENTS SHOULD BE APPLIED TO APPLICANTS WHO TRANSFER UBE SCORES AS ARE APPLIED TO THOSE WHO TEST LOCALLY.

this hampers score portability, so UBE jurisdictions follow a policy of notifying all applicants of their scaled scores.

POLICIES SET INDEPENDENTLY BY JURISIDICTIONS

All policies related to the requirements for admission on the basis of a transferred UBE score are left to the jurisdictions to set independently. As a general rule, when setting such policies, jurisdictions should keep firmly in mind that the UBE is the same exam, administered, graded, and scored uniformly by all UBE jurisdictions. Therefore, the same requirements should be applied to applicants who transfer UBE scores as are applied to those who test locally. There is no reason to differentiate on the basis of where applicants test.

Conditions for Accepting Scores

Setting Time Limits for Accepting Scores

Jurisdictions must decide how long a UBE score represents the applicant's current readiness to enter practice. NCBE does not make any recommendations in this regard, but NCBE Director of Testing Susan Case advises that jurisdictions should accept past scores for an interval that is reasonable to assume that the applicant's knowledge base has been maintained or has increased since the applicant took the exam.[21] In many cases, jurisdictions have already identified intervals for other issues, and those intervals might be equally applicable to transferred UBE scores. Any of the following could be used as a means of determining an appropriate time limit for accepting UBE scores:

- Jurisdictions typically set a time limit within which applicants who have passed the exam must complete the admission process and take the oath before their scores will be deemed stale and invalidated. A jurisdiction could apply the same interval to UBE scores transferred from other jurisdictions.

- If a jurisdiction accepts MBE scores earned in another jurisdiction as a basis for admitting applicants without further testing,[22] it might reasonably set the same time limit for accepting UBE scores as it applies to MBE scores.

- Similarly, jurisdictions that accept transferred MBE scores from prior examinations as a basis for allowing applicants to take only a portion of the current examination might apply the same time period to UBE scores. In this situation, however, arguments can be made for setting either a longer time limit for UBE scores (because the score represents the applicant's performance on the entire examination, not just the MBE) or a shorter time limit (because the applicant will not undergo further testing to assess current knowledge and skills).

Applying the Minimum Passing Score Consistently

Jurisdictions continue to set their own minimum passing scores and should apply the same cut-score standards to UBE scores transferred from other jurisdictions. NCBE recommends that UBE jurisdictions not condition acceptance of a transferred UBE score upon the applicant's passing status or admission to the bar in the testing jurisdiction. Doing so results in the receiving jurisdiction effectively adopting the minimum passing score of the testing jurisdiction in those cases where the testing jurisdiction's minimum score is higher; this causes different score requirements to be applied within the same jurisdiction. The following example illustrates the inconsistency that could result if this practice were to be followed.

> Assume that Jurisdiction A (the receiving jurisdiction) has a passing score of 260 and requires that applicants who transfer UBE scores must have passed in the testing jurisdiction (not recommended by NCBE).
>
> Applicant 1 earns a score of 270 in Jurisdiction B, where the passing score is 280.
>
> Applicant 2 earns a score of 270 in Jurisdiction C, where the passing score is 266.
>
> Applicant 1 could not qualify for admission in Jurisdiction A if Jurisdiction A were to condition acceptance upon the applicant's passing status in the testing jurisdiction. Although both Applicants 1 and 2 earn identical scores that exceed Jurisdiction A's minimum passing score of 260, the requirement that applicants must pass in the testing jurisdiction means that Applicant 1 must earn a score of 280 because that is the testing jurisdiction's minimum passing score.
>
> To take this illustration a step further, Applicant 3, who tests locally and earns a score of 260 in Jurisdiction A, could be admitted, while Applicant 1 with a score of 270 could not.
>
> Remember: *it's the same exam*, so a requirement that applicants must pass where they test should not be applied, because there is no common cut score.[23]

. . . [J]URISDICTIONS SHOULD ACCEPT PAST SCORES FOR AN INTERVAL THAT IS REASONABLE TO ASSUME THAT THE APPLICANT'S KNOWLEDGE BASE HAS BEEN MAINTAINED OR HAS INCREASED SINCE THE APPLICANT TOOK THE EXAM.

Legal Education Requirements

The same logic applies to legal education requirements. UBE jurisdictions should apply the same legal education requirements to applicants who transfer UBE scores as they apply to those who test locally. If a jurisdiction requires applicants to have graduated with a J.D. degree from an ABA-accredited law school to be eligible to sit for the examination, it should require the same of applicants who transfer UBE scores from other jurisdictions, even if the testing jurisdiction did not impose such a requirement. Presumably, jurisdictions require specific legal education as a prerequisite to sit for the examination because they believe that passing a bar exam, no matter how valid and reliable the exam, should not be the sole measure of preparedness to enter practice. Thus, jurisdictions should not alter their educational requirements for applicants who are transferring UBE scores merely because the applicants have already passed the examination.

Multistate Professional Responsibility Examination

If the jurisdiction requires exam applicants to pass the Multistate Professional Responsibility Examination (MPRE), it should require the same of applicants who transfer UBE scores. Applicants who are transferring UBE scores likely will also have taken the MPRE, since it is required by all but 4 of the 56 jurisdictions.[24]

But if a jurisdiction's rules require applicants to pass the MPRE within a specific time period relative to other events, such as within one year of passing the bar examination, the jurisdiction should consider whether to modify that requirement to coordinate with its conditions for accepting UBE scores. For example, if the jurisdiction accepts UBE scores earned within the preceding 24 months, many applicants transferring UBE scores may have to retake the MPRE to earn a more current score. For those applicants who are transferring UBE scores, jurisdictions might consider setting the time limit for passing the MPRE in relation to when the UBE score was earned.

It's a Score, Not a Status

Remember, it's the *score* that is portable, not the status. When developing a regulatory framework for accepting transferred UBE scores, jurisdictions should constantly return to the fact that *it's the same exam*. It doesn't matter where applicants test, just what scores they earn, and the requirements for admission should be consistent for those who test locally and those who test in other UBE jurisdictions.

As the UBE matures and is adopted by more jurisdictions, these policies may evolve to address new circumstances, and new policies may be developed. The key concepts of producing comparable scores, enhancing score portability, and ensuring reliable transfer of scores will continue to guide the process. ▨

Notes

1. Missouri and North Dakota administered the first UBE in February 2011. Alabama began administering the UBE in July 2011, while Idaho will start in February 2012 and Washington in July 2013. Use of the UBE is under consideration in other jurisdictions.

2. To learn more about the UBE, see the following *Bar Examiner* articles: Veryl Victoria Miles, *The Uniform Bar Examination: A Benefit to Law School Graduates*, The Bar Examiner, Aug. 2010, at 6; Susan M. Case, Ph.D., *The Uniform Bar Examination: What's In It for Me?*, The Bar Examiner, Feb. 2010, at 50; Susan M. Case, Ph.D., *Coming Together: The UBE*, The Bar Examiner, Aug. 2009, at 28; *Essays on a Uniform Bar Examination*, The Bar Examiner, Feb. 2009, at 6.

3. NCBE provides jurisdictions using any of its tests with an *MBE Supervisor's Manual*, *MEE Supervisor's Manual*, and/or *MPT Supervisor's Manual*, as appropriate, and it has now developed a *UBE Supervisor's Manual*.

4. UBE jurisdictions may choose to administer a jurisdiction-specific exam component in addition to the UBE to assess knowledge of local law, but the scores from any such component are not part of the portable UBE scores.

UBE Implementation:
Getting Started

When a jurisdiction is ready to adopt the UBE, it should review its rules for any changes that are necessary to conform to the UBE policies. It must also make policy decisions regarding those issues where uniformity with other UBE jurisdictions is not required, and decide what, if any, rule provisions are necessary to effectuate those policies.

In addition to amending its rules, there are other steps a jurisdiction should take to prepare to become a UBE jurisdiction. These fall roughly into two categories: (1) preparing to process a new category of applicants who are transferring UBE scores from other jurisdictions ("transfer applicants") and (2) preparing for changes to the existing examination to bring policies and practices into accord with the UBE.

NCBE is available to assist any jurisdiction by reviewing the jurisdiction's existing rules for conflicts with UBE policies and offering proposed language for any new rule provisions that might be necessary. In addition, NCBE can help identify some of the practical things that a jurisdiction might wish to do in preparation for becoming a UBE jurisdiction.

5. The two MPT tasks are administered in one session to save the time it takes to distribute and collect test materials in two separate sessions in larger jurisdictions.

6. A *banked score* is a score earned on one component in a prior examination in the testing jurisdiction, where the applicant did not pass the exam but scored high enough on one component so as not to have to retake that component. Allowing use of banked scores permits applicants to pass the exam in stages. A *transferred score* is a score earned in a prior examination in another jurisdiction, where the applicant may or may not have passed depending on the requirements set by the receiving jurisdiction for accepting transferred scores.

7. NCBE convened a group of bar admission administrators from nine jurisdictions to develop the model form. It is available to download from the secure section of NCBE's website that can be accessed only by administrators. Jurisdictions are advised to have the model form reviewed by their legal counsel before using it.

8. The grading workshop teaches graders how to calibrate but is not a calibration session by itself. Graders who attend the grading workshop should undertake additional calibration before beginning actual grading.

9. Use of abbreviations is permitted.

10. Score corrections due to mathematical error do not constitute regrading and are allowed. Such events are rare, however.

11. There are also practical reasons for not allowing post-release regrading by UBE jurisdictions. There is a period between release of results and the deadline for seeking regrading when an applicant might request that an official transcript be sent by NCBE to another jurisdiction. If the applicant subsequently petitions for regrading in the testing jurisdiction, there could be a difference between the transferred score and the "final" score.

12. NCBE's scaling services are offered free of charge to all jurisdictions, UBE and non-UBE.

13. For jurisdictions that administer a jurisdiction-specific exam contemporaneously with the UBE, NCBE will scale the jurisdiction-specific exam scores to the MBE so that the jurisdictions don't have to do any scaling calculations on their own.

14. Non-UBE jurisdictions may choose to receive their MBE scores rounded either to whole numbers or to one decimal. There are practical advantages to rounding scores to one decimal in that (1) consecutive raw scores do not result in the same scaled score when scaled scores are rounded to a decimal and (2) raw whole-number scores and scaled decimal scores are more readily distinguishable. *See* Michael T. Kane, Ph.D., *To Round or to Truncate? That Is the Question,* THE BAR EXAMINER, Nov. 2003, at 24, and Susan M. Case, Ph.D., *The Testing Column: MBE "Decimal Dust,"* THE BAR EXAMINER, Feb. 2004, at 33.

15. UBE jurisdictions may transfer their own MBE scores to non-UBE jurisdictions if they wish, but NCBE is the sole transferor of UBE scores.

16. NCBE makes no recommendation concerning whether jurisdictions should limit the number of times applicants may sit for the UBE; that is each jurisdiction's prerogative.

17. NCBE requires the applicant's name, date of birth, and Social Security number or NCBE number to identify UBE scores with the requisite degree of confidence jurisdictions should expect.

18. NCBE can suppress applicant names and Social Security numbers on the score roster sent to a jurisdiction if the jurisdiction's

policies mandate that it not receive this information with the MBE scores.

19. UBE jurisdictions that do not want the administrative burden of reporting scores to applicants can direct them to NCBE's website, where applicants may request an unofficial transcript of their scores.

20. Some non-UBE jurisdictions report scores only to failing applicants and not to successful applicants. In the context of the UBE, however, a score that is passing in the testing jurisdiction might not be passing in other jurisdictions. Hence, all UBE applicants are told their scores without regard to their pass/fail status in the testing jurisdiction.

21. A jurisdiction might decide to accept older UBE scores if the applicant has been engaged in the active practice of law for some portion of the time since the score was earned. In such circumstance, the jurisdiction is not relying solely upon the score (and completion of other admission requirements) as the measure of readiness to practice, but is coupling a passing score and experience practicing law.

22. North Dakota, Minnesota, and the District of Columbia admit applicants on the basis of an MBE score taken in another jurisdiction. *See* National Conference of Bar Examiners, Comprehensive Guide to Bar Admission Requirements 2011, Chart 8, 28–30 (National Conference of Bar Examiners and American Bar Association Section of Legal Education and Admissions to the Bar 2011).

23. In the context of accepting transferred MBE scores, many jurisdictions require applicants to have passed the examination in the testing jurisdiction, and some require applicants to have been admitted in the testing jurisdiction. Such a requirement makes sense for MBE scores because the transferred score is taken from only one component of the bar examination and the remainder of the examination is not uniform in all jurisdictions.

24. *See* National Conference of Bar Examiners, *supra* note 22.

Kellie R. Early is the Director of Administration for the National Conference of Bar Examiners. She was the Executive Director of the Missouri Board of Law Examiners when Missouri adopted the UBE.

EXHIBIT "C"

Judith A. Gundersen, *It's All Relative—MEE and MPT Grading, That Is*, THE BAR EXAMINER, June 2016

It's All Relative—
MEE and MPT Grading, That Is

by Judith A. Gundersen[1]

The Multistate Essay Examination (MEE) and the Multistate Performance Test (MPT) portions of the bar exam are graded by bar examiners in user jurisdictions. They are not centrally graded at NCBE, but NCBE prepares detailed grading materials for both exams and provides hands-on grader training designed to facilitate consistent, accurate, and fair grading across all MEE and MPT user jurisdictions and, in particular, for Uniform Bar Examination (UBE) jurisdictions. It is critical that grading of the MEE and MPT portions of the UBE be consistent across UBE jurisdictions, as the score portability afforded by the UBE is based on the assumption that the exams are graded in a consistent manner no matter where graded or by whom.

This article discusses the relative or rank-ordering grading philosophy NCBE uses in its grader training, the reasons we advocate this approach, and recommendations for optimal use of this grading method. I'll start with a review of how grading materials are prepared at NCBE and the nature of our grader training.

Preparation of MEE and MPT Grading Materials

Grading materials for the MEE and MPT include the MEE analyses and MPT point sheets, which are detailed discussions of all the issues raised in the items by the item drafters and suggested resolutions or analyses of the issues. The analyses and point sheets are drafted by the authors of the items and are then discussed, edited, and revised by the respective NCBE drafting committees at semiannual meetings. Preparing MEE and MPT items and their grading materials takes at least two years and is an iterative process with many lawyer-volunteers, NCBE staff, outside content experts, and pretesters involved.[2]

MEE and MPT drafters know the importance of crafting excellent grading materials. The process of preparing grading materials also serves as a good check for the drafting team on the item's internal consistency, degree of difficulty, and gradability; it is quite common for grading materials to unveil problems with the item that were not identified by the drafter or committee at an earlier stage.

Grader Training

MEE and MPT grading materials are very thorough and can effectively guide graders through the grading process. In addition, NCBE also conducts hands-on grader training sessions at its Madison, Wisconsin, headquarters the weekend following the bar exam. Graders may attend the grading workshop in person, by conference call, or via on-demand streaming as available following the workshop. Participation by user jurisdictions is high—hundreds of graders representing most MEE and MPT jurisdictions participate in one of these three ways.

The grading workshop lasts one day and consists of a dedicated session for each MEE and MPT item led by drafting committee members who are experienced grading workshop facilitators. Sessions begin with an overview of the item and grading materials, and any questions about the area of law (MEE) or the assigned task (MPT) are addressed. The participants then set about silently reading several real examinee answers (sent by bar administrators from all over the country) and grading them. Grades are assigned using a 1–6 relative score scale (as discussed later). As professors often do in law school, workshop facilitators rely on the Socratic method from time to time—graders are called on to explain the grades they gave. This is particularly true if a grade might be an outlier from grades assigned by other graders in the session. Based on the review and grading of the sample of examinee answers and the ensuing discussion between graders and facilitators, grading materials may be refined or grading weights adjusted. Final versions of the grading materials are then made available to graders in user jurisdictions a day or so after the workshop.

Grading workshop participation alerts graders to common answer trends and also gives them a head start on calibration—the development of coherent and identifiable grading judgments so that rank-ordering is consistent throughout the grading process and across multiple graders. (The focus of this article is *not* on calibration, but that doesn't mean it isn't a critical component of the grading process. See the section on calibration below.)

The Relative Grading Philosophy in Action

What Is Relative Grading and How Does It Work?

With NCBE's grading materials in hand, graders are ready to begin the grading process in their own jurisdictions with their own examinees' answers.

But grading MEEs and MPTs isn't like marking a paper with a score from 1% to 100% or meting out an A, B, C, D, or F (or drawing smiley or frown faces on papers; one of my sons' third-grade teachers, whom I will call Ms. Brinkman for purposes of this article, was fond of drawing a big ☹ on papers that didn't meet her standards!). Instead, NCBE trains bar examiners to grade the MEE and MPT on a relative basis—making distinctions between papers and rank-ordering them according to whatever score scale the jurisdiction has in place. (Jurisdictions may use whatever score scale they wish—e.g., 1–5, 1–6, 1–10, etc.—although NCBE uses a 1–6 score scale at its grading workshop, for reasons detailed later in this article.)

Relative grading training helps graders identify consistent standards in ranking papers and then apply those standards to put papers in piles according to their relative strength. The 1–6 scale used at the workshop simply means that a score of 6 is reserved for the best papers among all answers assigned to a particular grader. It is better than a 5, which is better than a 4, and so on, all the way to 1—a paper that is among the weakest papers. Relative grading means that in any group of answers, even if no single paper addresses all the points raised in an item, the strongest papers still deserve a 6 (using a 1–6 score scale). They do not have to be perfect nor necessarily deserve a true A or 100% (or a ☺☺ according to Ms. Brinkman). Using the same principles, a paper need not be completely devoid of content to get a 1 if the other papers are strong.

This relative grading philosophy (also referred to as norm-based grading) may be a little different from the way many of us had our papers graded in school, where we were held to an "absolute" or "criterion-referenced" standard: we had to answer a certain number of parts of a question correctly to get a high score or an A regardless of how our fellow

students answered. Or if we missed some points, we would get a low grade even if many of our fellow students also missed the same points.

NCBE's focus on relative grading does not mean, however, that absolute or criterion-referenced grading does not belong on the bar exam; it does, particularly on the Multistate Bar Examination (MBE)—the only part of the bar exam that is equated across time and across exam forms. Equating is the process of determining comparable scores on different exam forms. For the MBE, the absolute standard or "cut score" has the same meaning across administrations and jurisdictions. A scaled score (scaled means that it is a standardized score derived after equating) of 135 on the MBE is a 135 no matter when or where earned and will always mean that the examinee passes if the cut score is 135. By contrast, essays and performance tests cannot be equated in the way a multiple-choice exam like the MBE can be, so a total raw score of, say, 24 (or 100 or 1,000) on the written part of the bar exam may have a different meaning depending on the particular exam form, the examinee pool, the grader, and the jurisdiction.[3]

Because of the high-stakes nature of the bar exam, we must account for the differences in written exams across administrations, jurisdictions, and graders, and we do this by using the equated MBE score distribution as a highly reliable anchor. We weight the MEE and MPT raw scores for each examinee according to the jurisdiction's weighting scheme (e.g., on the UBE, the MEE is weighted 30% and the MPT 20%). We then map the total weighted MEE and MPT raw scores for each examinee to the MBE scaled score distribution according to performance level. This process is referred to as *scaling* and has the effect of adjusting the MEE and MPT scores so that they have the same mean and standard deviation as the MBE scores do in the testing jurisdiction (standard deviation being the measure of the spread of scores—that is, the average deviation of scores from the mean).

Scaling written scores to the MBE is a psychometrically valid practice because examinee performance on the MBE is strongly correlated to examinee performance on the combined MEE and MPT. Because the MBE is an equated exam, MBE scores have constant meaning across time and across jurisdictions, even though the items on particular exams may vary slightly in intrinsic difficulty. By scaling the combined MEE and MPT scores to the MBE scaled score distribution, we capitalize on (or leverage) the equating done to the MBE to give the MEE and MPT scores the same constancy in interpretation, despite the fact that MEE and MPT items may vary in difficulty from administration to administration.

It is important to point out that if the relative grading approach is used consistently across jurisdictions and administrations, the MEE and MPT raw scores will have the same mean and standard deviation in all jurisdictions and administrations no matter if the intrinsic difficulty of the MEE or MPT items changes or if the examinee population becomes more or less proficient. In jurisdictions that use the same grading scale, each jurisdiction will also have approximately the same raw score mean and standard deviation as well as having the same mean and standard deviation for all administrations. It is only by scaling to the MBE that differences in either the items or the examinees can be reflected in the scores.[4]

Why Use Relative Grading?

There are compelling psychometric and policy reasons why, given the current process for grading the MEE and MPT, NCBE trains graders to use a relative grading approach (with subsequent scaling to the MBE) to consistently grade the MEE and MPT.

Score Scales and Grading Procedures Vary Among Jurisdictions

Because of a decentralized approach to grading the MEE and MPT, no matter how successful we are at training graders across jurisdictions to promote uniformity, we must allow for the fact that there could be some scoring variation among jurisdictions.

Relative grading does not require that all jurisdictions use the same score scale. Rather, papers placed in a particular pile (assigned grade) reflect a level of proficiency that is more similar to others in the same pile than to papers placed in a different pile, and higher grades reflect higher degrees of proficiency. As stated earlier, an examinee's raw MEE and MPT scores are weighted appropriately, added together, and then mapped to the MBE scaled score distribution for the given jurisdiction. An examinee who performs well on all or most parts of the written portion of the exam will generally have scores that "land" on the upper end of the distribution of the MBE scaled scores for that jurisdiction. Someone who earns a lot of 6's on her MEE and MPT answers (in a jurisdiction using a 1–6 score scale) will generally have her total written score mapped to the top of the MBE scaled score distribution for her jurisdiction; an examinee who consistently earns 1's and 2's on his MEE and MPT answers will usually find that his total written score maps close to the bottom of the MBE scaled score distribution. This will be true no matter what score scale is used.

Relative grading is also adaptive enough to work with different approaches to the grading process. It does not matter if each paper is read by only one grader or by two graders who have to agree; or if a single grader grades all answers to a particular item or answers are divided among several graders. Nor does it matter if grading is done over the course of a day or weekend of intense grading or over the course of two months. As long as graders achieve

and maintain calibration (as discussed later), relative grading should serve to keep answer assessment consistent across time and across graders.

MEE and MPT Items May Vary in Difficulty from One Administration to the Next

As much as the drafting committees and our test development process try to standardize MEE and MPT difficulty across exam administrations, it is impossible to create items that represent exactly the same degree of difficulty. And MEE and MPT items cannot be pretested live to gather performance data in the way that MBE questions can because they're too few and too memorable. (MBE pretest questions are indistinguishable among the scored MBE items on each MBE exam form.) Without live pretesting, we must find some other fair way to take into account differences in MEE and MPT difficulty across exam forms.

With relative grading, it doesn't matter if an exam form represents the exact degree of difficulty as past (or future) MEEs or MPTs. Relative grading means that an examinee who sits for a harder exam is not penalized and an examinee who sits for an easier one is not rewarded, because it focuses only on how examinees do in comparison to one another on the same exam. For example, suppose that February 2016 examinees were given more difficult MEE or MPT items than those administered in, say, July 2015. That would be unfair to the February 2016 examinees or, alternatively, would seem like a windfall to the July 2015 examinees *if* MEE and MPT items were graded according to an absolute standard. The July 2015 examinees would get overall higher scores because the items were easier. In the world of high-stakes tests like the bar exam, this is a situation to avoid, and relative grading helps do that. It focuses on *comparing* answer quality according to other answers to the same items. Answers to easy items are still rank-ordered, as are answers to harder ones.

Scaling the total raw score on the written portion of the bar exam to the MBE, which is equated across administrations and accounts for differences in exam difficulty, means that it doesn't matter whether the written portion on one administration is harder than on another. As long as graders are able to rank-order answers, they can fairly and consistently grade the MEE and MPT from administration to administration regardless of differences in exam form difficulty.

Examinee Proficiency Varies from One Administration to the Next

Examinee proficiency may vary across administrations. For example, in the February administration, examinee proficiency tends to be lower due to a larger proportion of repeat test takers. We see this lower performance reflected on the MBE in February and expect to see lower scores on the MEE and MPT as well. However, asking graders to maintain consistent grading standards across administrations, examinees, and items would be extremely difficult, if not impossible. There are simply too many moving parts across test administrations to make such a grading task reasonable for maintaining score meaning across administrations. But relative grading—comparing answers among the current pool of examinees and then scaling those raw scores to the MBE—is manageable for graders and fair to examinees.

It is also important to note that using a relative grading system rather than an absolute grading system does not mean that graders are artificially inflating or deflating grades in a way that allows more examinees to pass or causes more examinees to fail. All relative grading does is help graders make rank-ordering decisions, which are critical to having the question "count" in an overall bar exam score, as discussed below. Scaling to the MBE lines up an examinee's overall written score to a statistically accurate corresponding point on the MBE score

distribution. Scaling standardizes rank-ordering decisions across time and exams.

Likewise, relative grading does not benefit or penalize examinees who sit in jurisdictions that have a weaker or stronger examinee pool. Relative grading practices work in tandem with the process of scaling to make the appropriate offset for each examinee's position relative to his or her own jurisdiction's examinee group and the position of that examinee group relative to other jurisdictions' examinee groups. To make meaningful and fair comparisons across time and jurisdictions, we need to know what absolute level of performance is represented by a particular group's average. We don't have that absolute performance information for essays, but we do have average performance on the MBE for the relevant groups. By virtue of the equating process, those scores are on an absolute scale.

Because the data have consistently shown across groups and time that the total MBE scaled score is strongly correlated with overall performance on the written components (correlation above .80 when reliability of the two measures is taken into account), we can use MBE performance information as a proxy indicator of the groups' general ability levels. As a result, an examinee whose total raw essay score is ranked at the top of a weak group will have, after scaling, a total scaled essay score that reflects that differential, and an examinee who is more toward the bottom of a strong group will have a total scaled essay score that accounts for that positioning as well. Similarly, offsets are made (via scaling) to account for an examinee who sits for an administration with easier essay questions or one who sits for an administration with harder essay questions. The scaling process is critical to ensure that scores have a consistent meaning and also to ameliorate any efforts at gaming the system by attempting to pick a group or an administration that

is anticipated to behave a certain way (e.g., sitting for a test that is anticipated to be easy or sitting with a group that is anticipated to be particularly skilled).

Graders Vary in Harshness or Leniency

In addition to evening out the differences in MEE and MPT difficulty from one administration to the next, relative grading ameliorates grader harshness or leniency from one administration to the next and from grader to grader. Even a harsh grader in jurisdiction A has to distribute an array of grades, low to high, among papers if she uses relative grading—she can't give all papers a low grade because then she's not rank-ordering. A lenient grader in jurisdiction B grading the same item can't give all papers a high grade if he uses relative grading and follows instructions to use all grade values.

If a particular question is graded by a harsh grader or a lenient grader, as long as that grader is consistently harsh or lenient in rank-ordering, examinees are not unfairly penalized or rewarded—the rank-ordering decisions made by the grader remain; the actual raw scores assigned, whether harsh or lenient, are smoothed out to fit the MBE scaled score distribution. Examinees will not be penalized even if harsher graders have a lower mean score than lenient graders. (Note that if multiple graders are assigned to grade a single question, they must be calibrated so that they do not have different levels of harshness or leniency.)

Relative Grading Facilitates the Equal Weighting of All Items

Relative grading facilitates spreading out scores, which is critical to ensuring that all items carry the weight they should in an examinee's overall written bar exam score. The weight an item gets is strongly affected by the amount of variation that scores have on that item. The less variation, the less weight the item carries in determining the total written score

value. A question that every examinee gets right doesn't discriminate or distinguish between examinees, just as a question that every examinee gets wrong doesn't discriminate. For example, a question asking an examinee to write the English alphabet wouldn't distinguish between examinees, because virtually everyone would get the answer correct. Or a question asking examinees to write the Burmese alphabet would probably stump 99% of U.S. examinees. In both instances, those questions would let us know that all examinees do know the English alphabet but don't know the Burmese alphabet, but they wouldn't provide any information to allow us to make distinctions *between* examinees based on their performance.

All MBE, MEE, and MPT items are designed to elicit information about examinee performance. Relative grading on the MEE and MPT, both of which have multiple issues per item, allows graders to gather information about examinee performance and assign a score that accurately reflects examinee performance. All MEE and MPT items are drafted, reviewed, edited, and pretested to ensure that graders will be able to spread examinee scores according to relative quality if they follow grading instructions properly.

Graders should award points or credit reflecting the spectrum of the score scale used in their jurisdiction to maximize and equalize the information provided by each MEE or MPT item. Consider the following examples of what happens when a grader fails to discriminate among answers. Suppose a jurisdiction uses a 10-point score scale and a grader is using an absolute or criterion-referenced approach (or, for that matter, a relative grading approach) and no examinees address all points raised in an item. The absolute grader won't award any papers a 10 or possibly even a 9, depending on how inadequate the answers are. And the relative grader, if not

trained properly, might hold back and not award maximum points even to the best answers. So now that 10-point scale (also being used to grade other items), for purposes of *this* item, is a de facto 8-point scale, because no one is getting a 9 or a 10. Suppose a grader is even more extreme and uses only 5 points on a 10-point score scale—from 2 to 6, for example. The result is that the item has even less impact on examinees' overall written scores in comparison to other items that are being graded on all points of the jurisdiction's 1–10 scale.[5]

Another way of not spreading out scores is by bunching a large percentage of scores in the middle of the score scale. For example, on a 1–6 score scale, a grader who gives 75% of her papers a 4 (whether using absolute or relative grading), is, in effect, downgrading the weight given to that item. This particular question has elicited very little information about examinee performance and has compressed the score scale from 6 points to just a few points—and mainly to one point, a 4.

One reason why we emphasize relative grading is that it should help graders spread out scores—no examinee has to write a perfect paper to get the highest score, and no examinee has to leave the page blank to get a very low score. As long as a grader keeps that principle in mind, it should be natural to spread out scores. And the actual score distribution for each grader is easy to keep track of and need not be in equal piles. It is enough to make meaningful distinctions between relative examinee performances that reflect all or most of any given score scale.

What Is the Best Approach for Optimizing Relative Grading?

Using a Manageable Score Scale

While relative grading works the same no matter the score scale, it tends to work *best* and is easiest to manage using score scales that are relatively compressed. For example, if a grader uses a 1–100 scale, it's conceivable that the grader could make 100 piles of rank-ordered answers, but 100 separate piles representing qualitative differences between answers can be pretty hard to wrap one's brain around. And, of course, a 1–100 scale brings to mind grading as it's done in school—absolute grading. That is, a 90–100 is an A and is reserved for an answer that covers all possible issues in the item, as opposed to an answer that is the best of a possibly weak group of answers. Also, probably for many jurisdictions that use a 1–100 scale, their graders assign grades by 10's—that is, 10, 20, 30, etc.—so that the score scale is really functioning more like a 10-point scale, not a 100-point scale.

NCBE uses a 1–6 scale to train graders, in part, because six piles of answers are manageable and memorable. And we use a 6-point scale instead of a 5-point scale because a 5-point scale resembles the A, B, C, D, and F grading paradigm that makes it a bit too easy to bunch scores on the midpoint or average—a 3 or a C. Using a 1–6 scale means that graders can't just label an answer average, or a 3—they have to make a decision as to whether it's a 3 or a 4, that is, a little bit above an average paper or a little bit below. Because many graders tend to bunch their answers in the middle (rather than at the ends of the score scale), just by using a 1–6 score scale rather than a 1–5 scale, they have to make a choice around that critical midpoint, which makes bunching harder and spreading easier. Some jurisdictions use a 1–10 scale, which also works well provided that all points on that score scale are awarded.

Participating in NCBE's Grader Training

While relative grading is fairly intuitive, it is informed and standardized by the MEE and MPT grading materials and the post-exam grader training that puts the relative grading principles into action.

That training emphasizes reading through several answers before assigning final grades to those first several papers read, as a grader won't yet have a good idea of what the overall pool of answers looks like. NCBE's grader training and materials also assign weights to subparts in a question. So an examinee who performs well on one subpart of an MEE question worth 25% of the total score that could be awarded for that question is not assured a 6 unless he performs well on the other parts of the question, too, in comparison with other examinees. In other words, there is a weighting framework for assigning points, which helps to keep graders calibrated and consistent.

NCBE also offers online support and hands-on workshops that demonstrate how to use relative grading. For most graders, it is not hard to rate answers in order of their relative quality. It might be a little more difficult to use the entire score scale, whatever that may be, but practices as simple as keeping track of the number of score piles and the number of answers in each score pile go a long way toward keeping score distribution (or lack thereof) front and center during the grading process. And consistency can be maintained by keeping benchmark papers for each score pile to illustrate what representative answers look like for each score—6's, 5's, 4's, etc. This is particularly important if grading is done over an extended period of time or by multiple graders.

Ensuring the Calibration of Graders

No grading process will be effective if graders (especially multiple graders assigned to a single item) are not calibrated. It should not matter to examinees who grades their papers or when their papers are graded. Relative grading and absolute grading both require calibration to be consistent and fair. Under either grading method, calibration requires reading through several sample answers,

a thorough understanding of and facility with the grading rubric, and agreement on the standards to be applied and how to apply them. In my experience, calibration is an exercise to which jurisdictions devote substantial resources, time, and verification to ensure that graders become and remain calibrated throughout the grading process.[6]

CONCLUSION

The relative grading approach is employed widely in the jurisdictions that administer the MEE and the MPT. When used in conjunction with scaling to the MBE and proper training and calibration of graders, relative grading promotes consistency and fairness in grading the written portion of the bar exam. It compensates for the use of varying score scales and grading procedures among jurisdictions as well as differences in the harshness or leniency of the graders themselves. It neither artificially inflates or deflates grades but facilitates the spreading out of scores, which are then scaled to the highly reliable anchor of the equated MBE score distribution. Finally, it ensures that any variation in difficulty of items from one administration to the next does not penalize or reward examinees, while facilitating the appropriate weighting of all items so that each item provides information about each examinee's performance. ▪

NOTES

1. NCBE Testing Department staff members Dr. Mark Albanese, Dr. Joanne Kane, Dr. Andrew Mroch, and Douglas Ripkey all provided assistance with this article.

2. For a detailed explanation of how MEE and MPT items and their grading materials are prepared, see my article in the June 2015 *Bar Examiner*: Judith A. Gundersen, *MEE and MPT Test Development: A Walk-Through from First Draft to Administration,* 84(2) THE BAR EXAMINER 29–34 (June 2015).

3. Some jurisdictions may grade on an absolute basis—awarding points according to the grading rubric—regardless of how other examinees answer the question. As long as this grading method is employed consistently and spreads out scores, it is an acceptable method of grading.

4. Precisely how the written portion of the exam is scaled to the MBE is complex; there is not enough space in this article to fully discuss it, nor am I qualified to explain how it is done from a true measurement perspective. For a complete discussion of the steps we undertake in scaling written scores to the MBE, see Susan M. Case, Ph.D., *The Testing Column: Demystifying Scaling to the MBE: How'd You Do That?*, 74(2) THE BAR EXAMINER 45–46 (May 2005); see also Mark A. Albanese, Ph.D., *The Testing Column: Scaling: It's Not Just for Fish or Mountains*, 83(4) THE BAR EXAMINER 50–56 (December 2014).

5. Minimizing a question's contribution to an overall written score is not necessarily problematic; if a question does not perform as intended, so that no examinees (or all examinees) get it right, then it is appropriate that that particular question's impact is minimal. Note that graders should not artificially spread scores just for the sake of spreading them. Distinctions made between papers should be material.

6. For a full discussion of calibration, see my Testing Column in the March 2015 *Bar Examiner*: Judith A. Gundersen, *The Testing Column: Essay Grading Fundamentals*, 84(1) THE BAR EXAMINER 54–56 (March 2015).

JUDITH A. GUNDERSEN is the Director of Test Operations for the National Conference of Bar Examiners. Among her responsibilities, she is program director for the Multistate Essay Examination and the Multistate Performance Test. Gundersen received her J.D. from the University of Wisconsin Law School.

EXHIBIT "D"

Mark A. Albanese, Ph.D., *The Testing Column: Regrading Essays and MPTs—And Other Things That Go Bump in the Night*, THE BAR EXAMINER, March 2016

THE TESTING COLUMN
REGRADING ESSAYS AND MPTs—AND OTHER THINGS THAT GO BUMP IN THE NIGHT

by Mark A. Albanese, Ph.D.

"To err is human," and bar examination graders are not immune to this maxim. It is probably not an overstatement to say that, on the first pass, most graders give the wrong grade to at least one paper in a given administration. In fact, graders in some jurisdictions may be grading under conditions that increase their risk of giving the wrong grade, particularly in jurisdictions in which graders may be sequentially grading hundreds or even thousands of papers under fairly severe time constraints.

Misgrading can be of many types. Some types are idiosyncratic to a given grader, whereby he or she gives grades that are higher (or lower) than other graders would give; this is known as rater bias. This can occur if a grader is more generous (or stingier) than the other graders in awarding points, or does not quite grade to the scoring rubric for an essay question and misapplies it consistently to all essays. Other types of misgrading may occur early in the grading process but work themselves out as the grader "gets in the groove." Some types of misgrading may occur more frequently as time passes, such as when fatigue starts to set in. Other types of misgrading may occur relatively randomly, such as from distractions that can occur at almost any time. Given that some incidence of misgrading is likely to occur during the grading of the written materials in almost all jurisdictions, what can be done about it?
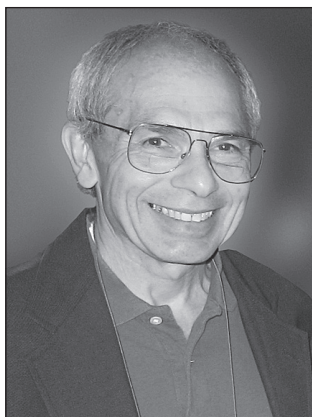
The best remedy is to avoid misgrading in the first place. The position of the NCBE Testing and Research Department is that grades on the individual written components (essays and performance test answers) are most accurate when provided by well-trained and calibrated graders who complete their evaluations in a time period that is as consolidated as possible. The primary strategy for achieving the goal of reliable grades contributing to valid score interpretations is grader training. Toward that end, NCBE provides grading workshops following each test administration. Jurisdictions are encouraged to participate in these workshops, which their graders can do either in person, by conference call, or via on-demand streaming after the workshop. Grading should also be done under conditions that are conducive to good concentration.

The next best remedy is to catch misgrading occurrences as they happen and correct them in real time. Essays that receive grades in the failing range can be submitted in real time to a second grader for a second look. This assumes that there are two graders for each essay, a practice that is not universal. Another option is to use MBE scores to predict examinees likely to be at the passing margin and have second graders who have been calibrated give a second look at each of the essays of those examinees.

The "least best" remedy is to regrade the papers of only those examinees who fail (an approach not recommended by NCBE). The problems with this approach and what can be done about them is the subject of the rest of this column.

Determining Who Is Eligible for Regrading

The most common regrading approach that jurisdictions choose is to regrade the papers of examinees who fail to pass. In a survey of jurisdictions conducted in April 2011 seeking information about regrading policies, to which 41 jurisdictions responded, 23 of the 24 jurisdictions that did some type of regrading only regraded those below the cut score. Some jurisdictions regrade all examinees who fail; others regrade only those within a certain range of the passing score. Psychometricians, however, being equal-opportunity analysts, recognize that there are two types of misgrading that influence whether an examinee passes. The first one is where examinees receive scores that are lower than they deserve. This is the type of misgrading that would be addressed by regrading only the failing examinees. However, the second type of misgrading is where examinees receive scores that are higher than they deserve. These are examinees who marginally pass but who would have failed had they been given the scores they deserve.

Depending upon your point of view, one or the other of these types of misgrading could be considered the more serious. From the examinees' point of view, an occurrence of misgrading that puts them on the fail side of the passing score is the more serious, since they must retake the bar exam, delaying their entry into the legal profession and, for many, increasing their already significant debt. From the perspective of the public, passing an examinee who should have failed is the more serious type of misgrading. After all, the whole licensing process

is in place to protect the public from incompetent lawyers. Bar examiners would also view this type of misgrading to be the more serious because their job is to protect the public from incompetent lawyers. However, the fear of lawsuits from failing examinees and sympathy for their plight often makes failing an examinee who should have passed the more serious problem for many in the bar examining community. The point is that if any papers are to be regraded, consideration should be given to regrading papers that fall both immediately above and below the passing score. The objective should be grading precision.

The Regrading Process Should Be as Fair and Unbiased as the Initial Grading Process

Bar examiners generally go to great lengths to ensure that the grading of examinee essays is as fair and unbiased as possible. Identities are replaced by codes to avoid any possibility of graders having any pre-existing bias against particular essays they are grading. Many jurisdictions put graders through rigorous training and calibration processes to ensure that the grading rubrics are being consistently and appropriately applied. Jurisdictions also employ quality-control measures to guard against "grader drift," the gradual shift of grading standards. Some of a grader's previously graded essays are embedded in his or her set of ungraded essays to make sure that the grader awards the same grade when looking at the same essay a second time. Something that is automatic during initial grading is that graders see the full range of answers to essay questions. This is important to help graders appropriately calibrate their application of scoring rubrics.

To summarize, in order for the regrading to be fair and unbiased, the following are essential: 1) essays must be de-identified, 2) graders must be trained to the same standards, 3) graders must be

calibrated to the same standards, 4) grader drift needs to be monitored, and 5) regrading should optimally occur above and below the preliminary pass/fail line.

1. Essays Must Be De-identified

If it were as simple as removing the examinee name and replacing it with a code, de-identifying essays for the regrading session would be a pretty straightforward process. However, in a regrading situation, it is difficult to avoid having graders know that the essays are being looked at a second time. Often, the graders are reassembled after a period of time has passed to specifically go through the regrading process. If the policy is to have all examinees who fail regraded, then graders will know going into the regrading process that they are looking at only the essays of examinees who failed in the first round of grading. Depending upon a grader's perspective, whether sympathetic or antipathetic to the failing examinee's plight, his or her grading may be biased. The only way to counter this bias is to intersperse the essays of failing examinees with the essays of a certain number of passing examinees and make this known to the graders.

2. Graders Must Be Trained to the Same Standards

After going through the process of grading the essays the first time, sometimes hundreds or even thousands on the same topic, bar examiners and graders may feel that graders have been adequately trained to grade a particular essay for the rest of their natural lives. However, just as time heals all wounds, it tends to make people forget painful experiences, and grading essays can be one of those. If a grading standard was set in the initial grading process—say, grading a standard set of 30 papers each to within one point on the grading scale—graders need to go through this process again before regrading to ensure that they are still applying the grading standard as intended.

3. Graders Must Be Calibrated to the Same Standards

Even if graders go through a set of papers during initial training to become calibrated, if all the papers they see during the regrading process are poor, it can begin to make papers that are less poor actually look good. This tends to distort the regrading process because the really good answers are not there to provide adequate context for applying the grading rubric. This distortion can be avoided during regrading by including papers that were given the full range of grades in the set of those being regraded. As noted above, this will also avoid rater bias that can occur if only failing papers are regraded.

4. Grader Drift Needs to Be Monitored

Grader drift needs to be monitored if each grader is regrading more than about 20 papers. There is no fixed point where drift is known to occur, but cognitive scientists have consistently found that short-term memory can hold seven, plus or minus two, unrelated things.[1] So, assuming that the same dynamic holds for remembering essays and the grades awarded, graders can remember essays and the grades they award for somewhere between five and nine essays. Beyond that point, each new paper graded replaces one graded earlier from short-term memory. The more papers graded, the further the papers in short-term memory separate from the first paper graded, and the grades awarded can drift. Interspersing a previously graded essay after approximately 20 "fresh" papers will enable drift to be monitored.

5. Regrading Should Optimally Occur Above and Below the Preliminary Pass/Fail Line

If precision is the goal, and it should be, graders need to have an adequate range of papers to grade such that they can be properly calibrated and unbiased in their grading. Therefore, regrading should include

an equal number of papers from above and below the pass/fail line.

Summary Remarks

Regrading is not a subject that has received much attention. Some jurisdictions engage in some form of regrading; others do not. Regrading recently came to the fore in discussions with several jurisdictions during which it became clear that some are strongly wedded to regrading failing examinees, even to the point of having a direct mandate to do so from their Courts. As a consequence, NCBE technical staff were compelled to offer their recommendations.

NCBE does not recommend regrading any part of the bar examination. We strongly recommend that jurisdictions put their entire grading resources into making certain that the initial grades awarded are of the highest possible caliber. We put our money where our mouth is and invest substantial resources and effort into providing grading workshops after each bar examination. Best practice, and what is probably the industry standard, is to have two graders grade each essay. This may not be practical for many jurisdictions.

Looking to the horizon, an alternative that is being used more frequently in other forms of testing is to employ computer grading as a second grader. Artificial intelligence grading applications are becoming increasingly sophisticated, to the point where the results are as reliable as those of human graders.[2] Whether computer-generated grades are as valid is a point where the jury is still out, but on the Certified Public Accountant (CPA) exam,

responses to the written communications section are scored by a computer grading program, and human graders are only being used if a score is close to the passing score. So, ready or not, computer grading applications are being used operationally by at least one licensing test.[3] NCBE researchers are currently investigating the state of the art in the realm of computer grading for assessing legal writing, and we will have a report on this that may become fodder for a future Testing Column.

In closing, there are technical aspects to deciding whether or not to regrade essays and MPTs and, if so, how to go about it, and I have put forth the technical recommendations of the NCBE Testing and Research Department in this column. But there are political and policy issues that may be as significant or more significant. It will be important to address these issues as we go forward, even if it is to agree to disagree. The practice of regrading essays and MPTs may still go bump, but it should no longer be at night. 🔳

Notes

1. G.A. Miller, *The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information,* 63 Psychological Review 81–97 (1956).

2. M.D. Shermis & B. Hammer, "Contrasting State-of-the-Art Automated Scoring of Essays: Analysis" (presentation, Annual Meeting of the National Council on Measurement in Education, Vancouver, British Columbia, April 16, 2012).

3. American Institute of CPAs, How Is the CPA Exam Scored?, www.aicpa.org/BecomeACPA/CPAExam/PsychometricsandScoring/ (last visited Feb. 4, 2016).

Mark A. Albanese, Ph.D., is the Director of Testing and Research for the National Conference of Bar Examiners.