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. 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. 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. 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**


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