

# THE TESTING COLUMN

## PROCEDURE FOR GRADING ESSAYS AND PERFORMANCE TESTS

by Susan M. Case, Ph.D.

**G**radings essay and performance test (PT) answers is one of the most important processes in bar examining. This article proposes a procedure for grading that will help ensure the quality of this process. The procedure outlined below includes tasks for staff as well as for the graders and begins shortly after test administration. It does not include the myriad of tasks that are undertaken before test administration occurs.



information regarding repeater status, test site, school, or anything else, each paper should be assigned a non-meaningful ID number. Jurisdictions that assign ID numbers to examinees that indicate repeater status or testing site should assign ID numbers to the papers that mask these distinctions.

**Distributing papers to graders.** After copies are made or scans are completed,

staff members distribute the examinee answers to the graders. If the answers to each essay question and PT are divided among multiple graders, efforts need to be made to ensure that the quality of batches of examinee answers is equivalent. For example, staff members should divide the papers so that each grader has the same number of papers from repeaters and from first-time takers (i.e., half the repeater papers go to grader A and half to grader B if the papers are divided between two graders). Also, it is a good idea to divide papers by school so that each grader has the same number of papers from each school.

### STAFF WORK BEFORE GRADING BEGINS

**Making copies of examinee answers.** We recommend that all examinee answers be scanned to create an office file of answers and that a copy be used for grading. Today's scanners are relatively inexpensive, and the scanning process requires no more staff time than photocopying. Unfortunately, there have been several incidents in which examinee answers were lost in the grading process and copies had not been made. This problem can readily be avoided by retaining a copy of all examinee answers in the office.

**Assigning ID numbers to papers.** An important staff task is to organize the grading process to ensure that the grading is done fairly, efficiently, and accurately. To make sure that the papers are completely anonymous, with no identifying

**Preparing for the calibration session.** In advance of the calibration session, it is generally preferable to send each grader the essay question or PT, the grading guidelines, and a few examinee responses to allow each grader to come fully prepared to the calibration session. Staff members should select at

least 30 papers for use in calibration. These papers should represent a cross section of quality—some from first-time takers, some from repeaters, some from each school. Selection of typed or neatly written papers will make the job of calibration easier; as long as these papers represent the range of quality, the selection of calibration papers may be made from among the typed papers with at most a small number of handwritten papers. Copies of the calibration papers will need to be made so that each grader has a copy. These calibration papers will also be used during the grading itself.

**Conducting the calibration session.** The calibration session should begin with an orientation for all graders. This should include a general discussion about the purpose of the written component of the bar exam, and should clarify the extent to which attributes such as writing style, organization of the written response, grammar, and spelling affect the grade. There should be rules regarding the use of zeros, instructions on how to grade a blank paper, and instructions on how to grade a paper that is off point (e.g., a diatribe on the bar exam instead of an answer to the question). Even if these issues have not arisen to date, it is best to have procedures in place to deal with them.

Graders also need to understand that as long as grades are scaled to the MBE (as they are in most jurisdictions), grading papers stringently or leniently will not affect the number of examinees who pass or fail. The grades rank-order the examinees and affect which of them are higher or lower on each question, but the cut score and the number of passes are determined by the MBE scores for the group of examinees. For example, if the average grade for a group of examinees on the written portion of the

test is 40 and the average MBE score for that group is 140, the 40 will be converted to a 140 in the scaling process. If the graders are more lenient and the average grade is 45, the 45 will also be converted to a 140 (since the MBE score for the group does not change). Similarly, if the graders are more stringent and the average grade is 35, the 35 will be converted to a 140, again because the average MBE score for the group does not change. So graders should think about rank-ordering the candidates on the essay or PT they are grading and should spread scores out as much as possible, using every score point.

## GRADER WORK BEFORE THE CALIBRATION SESSION

Before the grading begins, graders should review the question that they are grading as well as the grading guidelines. Graders should do any research necessary to ensure that they are completely familiar with the legal issues in the question. It is also helpful for graders to review the sample of examinee answers before the calibration session.

Graders might find it useful to expand the grading scale by using pluses and minuses to indicate those papers that are believed to be especially strong or weak representatives of their grading “buckets.” For example, a 5+ would indicate a paper that is better than most 5s but not quite good enough to be a 6.

## PROCEDURE FOR JURISDICTIONS THAT HAVE ONE GRADER PER QUESTION

- Read 5 papers, rank-order them, and assign grades.
- Read another 5 papers, rank-order them, and assign grades.

- Read another 10 papers, either in batches of 5 or all together, and assign grades.
- If you have not used all the grading buckets, grade another 5 papers.
- If you have still not used all the grading buckets, rework the grading criteria to achieve a spread of grades across the score scale. Grade another 5 papers to ensure that the changes are appropriate.
- Sort the papers according to the grade assigned and reread the papers for each grade to ensure that the quality of the papers assigned the same grade is comparable.

#### PROCEDURE FOR JURISDICTIONS THAT HAVE MULTIPLE GRADERS PER QUESTION


- Follow the procedures outlined above, but in addition, after each set of 5 papers, compare grades with your team and discuss any that are not matches.
- Clarify any of the grading guidelines as necessary to ensure consistency across the graders.
- Continue to follow this process with additional sets of 5 papers until consistency among the graders is achieved.

#### STAFF WORK AFTER THE CALIBRATION SESSION

**Monitoring calibration.** Grading should be monitored to ensure that the calibration remains in place. For every 100 papers that are being graded, approximately 10 of the calibration papers should be inserted. Regrading the original papers in stream

with the new papers will help ensure that the use of the grading scale remains consistent. Some graders like to retain some sample calibration papers from each grading bucket to serve as benchmarks and refer to these as they grade. Graders should return grades centrally every day (or every half day if they are grading full-time as in a grading retreat). Staff members should note grades of calibration papers and notify graders of any discrepancies from the grades assigned during the calibration session. In rare instances, there might be enough grader drift to warrant a repeat calibration session, but usually a conversation among graders is all that is required. It is best to conduct the grading process as quickly as possible after the calibration session to ensure that graders do not drift over time.

**Scaling written grades to the MBE.** NCBE will scale your written grades to the MBE at no cost. Turnaround is very quick, generally within a day. The scaling formula should be generated using only bona fide examinees who took the entire test; those who carried over a test score should not be included. Anyone who left the test without completing it should also be excluded. The scaling formula may then be applied to anyone who carried over a test score from a previous administration. If you send your scores to NCBE for scaling, please note which examinees took the entire test.

Some jurisdictions may have special issues that are not covered in this article. If so, please contact us and we will work with you to ensure that your grading is fair to the public and to the examinees. 

SUSAN M. CASE, PH.D., is the Director of Testing for the National Conference of Bar Examiners.