

THE TESTING COLUMN

TOP 10 LIST OF BEST PRACTICES IN TESTING FOR ADMISSION TO THE BAR

by Susan M. Case, Ph.D.

At our annual conference held this year in Austin, I was asked to speak on the topic “Ten Best Practices in Testing.” This column revisits that talk.

Wikipedia provides a good definition of *best practice*:

A best practice is a technique, method, process, . . . that is believed to be more effective at delivering a particular outcome than any other technique, method, process, etc. when applied to a particular condition or circumstance. The idea is that with proper processes, checks, and testing, a desired outcome can be delivered with fewer problems and unforeseen complications.

The term “best practices” has implications of finality, obedience, authority, and universality. (*Wikipedia*, “Best practice,” http://en.wikipedia.org/wiki/Best_practice [last visited April 28, 2010])

Best practices refers to more than a collection of good ideas; the term *best practice* implies that the practice has been investigated and carefully considered. Often it implies that the practice is designated as “best” as the result of research into alternative practices.



High-stakes testing demands that every effort be made to use best practices because the methods used affect the validity of test scores. Everyone involved in bar admissions testing has a responsibility to ensure that the test scores reflect what each jurisdiction has included in its definition of “fitness to practice,” and that the scores are accurate and reliable enough to be fair to both the examinees and the

public. If the exam scores are affected by something unintended (by attributes that are unrelated to fitness to practice), these scores will not produce the desired effect of identifying those who have sufficient knowledge and skills to begin practice as newly licensed lawyers.

The *Standards for Educational and Psychological Testing*, developed jointly by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, states the following:

Tests used in credentialing [such as the tests used in bar admissions] are intended to provide the public . . . with a dependable mechanism for identifying practitioners who have met particular standards. . . . Credentialing also serves to protect the profession by excluding persons who are deemed to be not qualified to do the work

of the occupation. Tests used in credentialing are designed to determine whether the essential knowledge and skills of a specified domain have been mastered by the candidate. (AERA, 1999, p. 156)

Validity can be enhanced by avoiding anything that would inadvertently affect the meaning of an examinee's score, and by ensuring that all aspects of the examination process, including test development, test administration, grading, equating, scaling, score combining, and score reporting, are of the highest quality.

My top 10 list of best practices in testing includes the following.

1. Each exam component must have a stated purpose.

The first step in developing a testing program is to create an explicit statement of the purpose of each test that indicates why each test is an important part of the bar admissions process. The purpose statement should be public information and should be clearly understood by those writing, reviewing, and grading test questions.

Each test must also have sample questions and a content outline that delineates what the test will include and, by exclusion, what will not be included.

For example, the purpose of the MBE, as stated in the *MBE Information Booklet*, is to assess the extent to which an examinee can apply fundamental legal principles and legal reasoning to analyze given fact patterns. This statement tells all interested parties that the MBE focuses on application of knowledge and that merely memorizing facts is not enough to succeed on this exam. Examinees must have learned the facts but must also be able to analyze the facts, decide which facts are applicable, and apply those facts to the cases that are presented in the questions. The purpose statement and the 18 sample questions

in the booklet are designed to provide a clear indication of what to expect on the MBE.

2. Each exam component must be developed using professional standards of test development and with strictest adherence to security.

It is important to ensure that the item authors are expert in the topics they are writing on. To ensure complete accuracy in the drafting and grading process, it is not appropriate to rely exclusively on those with only general knowledge of a topic. Each question and its grading material should be reviewed by at least one (and preferably more than one) content expert beyond the person who wrote the question.

Each question should also be reviewed to ensure that it is consistent with the exam purpose and covers material that is important. For example, to assess the extent to which each MBE question is relevant to the practice of a newly licensed lawyer, each question is rated by six pairs of reviewers (two reviewers per content area) who are not otherwise involved in the test development process. Their feedback is returned to the committees that write the MBE questions in time for them to review, alter, and discard questions as appropriate. For the last exam reviewed, 91% of the questions were rated highly or moderately relevant to the practice of a newly licensed lawyer; questions were then modified as necessary before the final exam was compiled.

Professional test development standards require that each question be completely clear. Examinees should not have to wonder what is really being asked. Repeated review helps to ensure question clarity, but reviews cannot accomplish everything. Pretesting is also important to ensure that each question is well constructed and appropriate in terms of difficulty. All multistate test questions are pretested, and pretest data are reviewed by the drafting committees in time for them to alter or discard questions as appropriate.

Finally, security must be maintained throughout the test development process. Test developers and users should look for the weak links in the process and carefully consider who has access to the actual questions, the grading materials, and even the topics being tested.

3. Grading criteria must reflect the exam purpose, and the grading processes must adhere to professional standards.

This practice seems obvious, but the challenge is in the details. Grading should be consistent with the purpose of the test. If the stated purpose of an exam component is to assess the extent to which an examinee can compose a well-written response, then the grading guidelines must take this purpose into account. For example, grading guidelines should be clear on the extent to which spelling, grammar, writing style, and following directions count.

4. Each exam question must be reviewed and pre-tested to ensure the quality of the test development and grading criteria.

Each question should undergo review by at least three different groups: content experts who review each question and the grading guidelines, newly licensed lawyers who answer the questions under exam conditions, and staff who review the questions for editorial style and consistency with other questions.

The only way to ensure that test questions are appropriate in terms of difficulty and that the grading materials appropriately distinguish between higher-quality and lower-quality papers is to pre-test the test and grade the answers using the grading guidelines. Pretest results will help to inform decisions about the grading guidelines. For example, if the pretesters do not perform as well as expected on a question, or if they miss a point that they were expected to include in their responses, the test developer should review the directions, the question, and the grading guidelines to make sure the question is clear.

5. Test administration practices must ensure that each examinee is authorized to take the test.

It is important to verify the identity of those sitting for the tests. Those administering the tests must check IDs carefully and take steps to ensure that each examinee fills in the proper ID on his or her answer sheet. ID fraud takes place when a ringer, or surrogate, takes the test for the examinee using a fake ID, or when a ringer takes the test and turns it in as the examinee's paper, while the examinee turns in a paper for the ringer. (Obviously the ringer would be someone who is not legitimately seeking licensure in that jurisdiction.) Proper ID checks, careful proctoring, and not allowing someone to take the exam who has previously passed it would reduce the likelihood of a ringer taking the exam and help prevent ID fraud.

6. Test administration practices must ensure that examinees do not have access to testing aids.

The Internet describes many possible illicit testing aids, including old standbys such as notes written on body parts and clothing. Newer methods involve notes written on the inside of a bottled water label or a candy wrapper.

7. Test administration practices must ensure that examinees cannot copy from one another.

The issue of copying was studied at NCBE by asking appropriately aged examinees to try to copy from one another's papers. Based on this exercise, it is recommended that a single examinee be seated at a four-foot or six-foot table. If two examinees sit at a single table, eight-foot tables should be used. Volunteer cheaters were able to copy from one another's papers when seated less than eight feet apart. Seating should not be staggered; examinees should be seated in a line from the front of the room to the back of the room and should not face each other across the table. Examinees should be required to sit at assigned seats and should not be allowed to sit where they prefer.

8. Test administration practices must ensure that examinees cannot take test material or information out of the testing room.

Proctors need to be vigilant to ensure that examinees do not remove test material from the testing

room, either by removing pages from the test book or by using a recording device, such as a pen or highlighter with a built-in camera.

9. Grading practices must follow professional standards, with emphasis on grader training, calibration, grading consistency, and monitoring.

It is important to ensure that each grader understands the purpose of the test, remains calibrated throughout the grading process, and uses the whole score scale. If more than one grader is used per essay, the multiple graders should be calibrated to one another, and checks should be made to ensure that they remain calibrated throughout the grading process.

If the essay grades are scaled to the MBE, there is no disadvantage in using the whole score scale. If no one does an outstanding job on the essay, the highest grade should be awarded to the best papers, even though the best papers failed to achieve perfection. Similarly, the worst papers should be assigned the lowest grades. Note that if essay grades are scaled to the MBE, assigning only high grades (or assigning only low grades) does not affect the passing rate. Grades should be spread out to the extent possible, and should reflect the rank order of papers where papers with higher grades are of higher quality than papers with lower grades.


For further information about how to calibrate graders, see “Practices and Procedures to Improve Grading Reliability on Essay Examinations: A

Guide to the Care and Feeding of Graders,” by Kellie R. Early, in the November 2004 issue of the *Bar Examiner* (available at http://www.ncbex.org/uploads/user_docrepos/730404_early.pdf).

10. Scores must be equated, scaled, and weighted to ensure appropriate score meaning.

In order to ensure consistent standards and fairness to all examinees, test scores must be equated across time. This ensures that any differences in exam difficulty are evened out and that the test scores retain the same meaning over time.

Equating can only be done directly with multiple-choice tests. The only way to equate essay scores is by scaling them to the MBE. Weighting the MBE at least 50% ensures that the reliability of the total score is at least 0.90, the minimum standard for high-stakes tests.

This column outlines 10 best practices in testing, but there are many more that could have been included in this column. Best practices is one of the areas in which NCBE does a great deal of research, with the goal of always improving the quality of the bar admissions process. As results of NCBE’s research become available, they will continue to be presented in the *Bar Examiner*. 

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