

# THE TESTING COLUMN

## SOMETIMES BOGSAT IS JUST NOT ENOUGH

by Susan M. Case, Ph.D.

I am constantly surprised by the new things I learn only to find out that everyone else already knows them. Readers of this column may remember that last year I traveled to the United Kingdom to help develop questions for a new exam program in England and Wales to test foreign-trained solicitors.<sup>1</sup> I returned to the UK in June of this year to again assist with question development. While discussing various issues, the group referred to some decisions being made by BOGSAT, a term that was unknown to me.



I learned that the acronym stands for Bunch of Guys Sitting at a Table. A Google search showed 13,100 results. Bing Dictionary defines BOGSAT as “decision-making by committee: the management practice of using often inexperienced committee members to make the most important decisions.”<sup>2</sup> The term was coined during the Kennedy administration in September 1961 by Waldemar Nielsen, an American author and expert on philanthropy, as a humorous explanation of how decisions were being made.

I endorse the use of BOGSAT for some purposes; it is an efficient way of making many decisions, especially if the “guys” around the table are an appropriately diverse group of knowledgeable individuals who are widely representative of the relevant points of view and areas of expertise. However, there are times when BOGSAT is just not good enough, par-

ticularly when the situation requires data for decision making.

Three topics that are frequently encountered in bar admissions are described below. Decisions on each have historically relied on BOGSAT, even though this process fails to meet professional standards for high-stakes decision making, as defined in the *Standards for Educational and Psychological Testing*. The *Standards*—

developed by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education—provide criteria for the evaluation of tests and are generally accepted by measurement experts.

### Exam Content Coverage

The *Standards* require that those who are developing exams for licensure (known as “credentialing tests”) develop content specifications based on a rationale for why the listed knowledge and skills are being included on the exam. As is noted in the excerpt below (see especially the Comment), the method for deciding what should be included is a job or practice analysis (a study of what newly licensed practitioners are required to do to perform their jobs competently).

#### Standard 14.14

**The content domain to be covered by a credentialing test should be defined clearly and**

**justified in terms of the importance of the content for credential-worthy performance in an occupation or profession. A rationale should be provided to support a claim that the knowledge or skills being assessed are required for credential-worthy performance in an occupation and are consistent with the purpose for which the licensing or certification program was instituted.**

*Comment:* Some form of job or practice analysis provides the primary basis for defining the content domain. If the same examination is used in the licensure or certification of people employed in a variety of settings and specialties, a number of different job settings may need to be analyzed. Although the job analysis techniques may be similar to those used in employment testing, the emphasis for licensure is limited appropriately to knowledge and skills necessary for effective practice.... [In] tests used for licensure, skills that may be important to success but are not directly related to the purpose of licensure (e.g., protecting the public) should not be included. For example, in real estate, marketing skills may be important for success as a broker, and assessment of these skills might have utility for agencies selecting brokers for employment. However, lack of these skills may not present a threat to the public and would appropriately be excluded from consideration for a licensing examination. The fact that successful practitioners possess certain knowledge or skills is relevant but not persuasive. Such information needs to be coupled with an analysis of the purpose of a licensing program and the reasons that the knowledge or skill is required in an occupation or profession.<sup>3</sup>

Although NCBE has used test development and reviewer groups that are far broader and more diverse than are possible with state-developed com-

ponents, we recently took the important step of conducting a full-scale job analysis to investigate what newly licensed lawyers actually do in their practices and what knowledge, skills, and abilities this requires. Information from this analysis will be used to evaluate our existing exams and to decide if additional content areas should be tested or if additional skills should be assessed.

In order to satisfy Standard 14.14, the same sort of job analysis process should be undertaken for any state component.

### **Score Reliability**

The *Standards* require that reliability be determined for any scores on which decisions are based.

#### **Standard 14.15**

**Estimates of the reliability of test-based credentialing decisions should be provided.**

*Comment:* The standards for decision reliability . . . are applicable to tests used for licensure and certification. Other types of reliability estimates and associated standard errors of measurement may also be useful, but the reliability of the decision of whether or not to certify is of primary importance.<sup>4</sup>

NCBE routinely analyzes the reliability of the scores from the MPRE and the MBE. In addition, NCBE has analyzed the scores from written components when jurisdictions have requested it. These analyses have led us to strongly recommend that the pass/fail decision be based solely on a total score and not based on achieving a passing score on each component of the exam. We have not found that scores based on these separate components have sufficient reliability for decision making.

The issue of score reliability cannot be decided by committee; reliability calculations require analyses by a psychometrician, followed by discussions

about exam construction to ensure that reliability levels are high enough to meet professional standards. NCBE will analyze the data from any jurisdiction upon request at no cost; however, the analysis needs to be followed by modifications to the exam program if the reliability is not sufficiently high.

### Passing Standards

The *Standards* require that policy makers determine an appropriate passing standard.


#### Standard 14.17

**The level of performance required for passing a credentialing test should depend on the knowledge and skills necessary for acceptable performance in the occupation or profession and should not be adjusted to regulate the number or proportion of persons passing the test.**

*Comment:* The number or proportion of persons granted credentials should be adjusted, if necessary, on some basis other than modifications to either the passing score or the passing level. The cut score should be determined by a careful analysis and judgment of acceptable performance....<sup>5</sup>

Procedures have been developed for setting an appropriate pass/fail standard for an examination used for high-stakes decision making. These procedures are followed by every licensing and certification exam that I am aware of *except* the bar examination. For example, the National Board of Medical Examiners, which develops the exams used to license physicians, reviews the pass/fail standard for each exam every year and conducts full standard-setting studies every three years on each exam. The passing standard is frequently modified based on the data from the standard-setting studies. Similar processes are followed by every medical specialty board, and by the accountants, the architects, the engineers, the nurses, the physical therapists, and so on. These studies can be resource-intensive, but they

are a responsibility of those who use high-stakes tests for decision making.

The three practices described above (determining what should be covered on the examination, ensuring that the scores' reliability is adequate for licensure decisions, and setting an appropriate pass/fail standard) are critical components of developing a high-stakes testing program that is used for deciding who can and who cannot be licensed to practice law. The arguments are often made that the costs of adhering to these standards are too high, that law students (to whom these costs would be passed along) already have enough debt, and that the budget just doesn't allow for this use of resources. However, the duty of bar admissions is to protect the public, and this requires that the tests are as good as they can be—covering the most important knowledge and skills, generating scores that would reliably place the candidate in the passing group or the failing group whether he or she took this exam or the next one, and drawing the pass/fail line at the appropriate place. The economic argument can be viewed from the other side: after spending significant money on a law school education, doesn't every applicant deserve the highest-quality bar exam that meets professional standards? 

## NOTES

1. For details of the exam program, the Qualified Lawyers Transfer Scheme, see Susan M. Case, Ph.D., *Testing Foreign-Trained Lawyers: A New Exam in England and Wales*, 80(3) THE BAR EXAMINER 34–37 (Sep. 2011), available at [http://www.ncbex.org/assets/media\\_files/Bar-Examiner/articles/2011/800311TestingColumn.pdf](http://www.ncbex.org/assets/media_files/Bar-Examiner/articles/2011/800311TestingColumn.pdf).
2. *Bing Definition*, s.v. “BOGSAT,” <http://www.bing.com/Dictionary/search?q=BOGSAT&q&s=n&form=QB> (last visited July 31, 2012).
3. AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, AMERICAN PSYCHOLOGICAL ASSOCIATION, AND NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION, STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING 161 (American Educational Research Association 1999).
4. *Id.* at 162.
5. *Id.*

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