

Garvey, a highly experienced New Hampshire lawyer and chair of his firm's trial department, as a professor of law and director of the Daniel Webster Scholar Program. Garvey is now charged with working with members of the Webster Scholar Committee and Pierce Law faculty to design and implement the program, including the practice courses and logistics of the comprehensive evaluations.

In addition to hiring a director, one other task has already been completed. After going through its required rulemaking process, the New Hampshire Supreme Court amended its rules to allow for bar applicants to be admitted to the practice of law after they have successfully completed the Webster Scholar Program. The Webster Scholar bar applicants will still be required to pass the MPRE and the New Hampshire character and fitness requirements. The court's rule amendment took effect on July 1, 2005.

Recognizing the difficulties of launching a project of this scope, the Webster Scholar Committee decided to implement the program as a three-year pilot program, with the hope that Pierce Law Center would fully fund and make the program available to many law students at the end of the three-year pilot period. During the pilot phase, the Webster Scholar Committee plans to continue meeting monthly with the director, providing guidance and feedback about the program's development.

There is an advantage of creating this variant to the bar examination in a state like New Hampshire, a state with relatively few attorneys and only one law school. Under these circumstances, it is much easier to regularly engage in conversations with judges, lawyers, and licensing officials. As David Leach, M.D., noted during his presentation at the October 2004 Joint Working Group Conference co-sponsored by the AALS, NCBE, and the ABA, the quality of

what we do "is directly related to the quality of the conversations in our lives." Here in New Hampshire, we have had monthly, documented conversations, enabling us to build upon and improve the quality of those conversations.^{2,3} 

ENDNOTES

1. The program is named after Daniel Webster, one of New Hampshire's most distinguished lawyers.
2. We also recognize that a number of other factors have enabled us to move from idea to implementation. Members of the committee know and respect each other; each is committed to make this program successful. Our chair, Justice Dalianis, has led the process, setting rigorous agendas and marshalling resources. Her colleague, Justice James E. Duggan, is a former law professor and acting dean at Pierce Law; he understands the issues involved in creating such a program from many angles. The chair of the New Hampshire Board of Bar Examiners, Frederick J. Coolbroth, is interested in other ways to examine lawyers. Former New Hampshire Bar Presidents Bruce W. Felmy and Martha Van Oot are leading attorneys in the state, and knowledgeable about legal education. Attorney Lawrence A. Vogelman is a member of the New Hampshire Board of Bar Examiners and is a former clinical professor. Pierce Law Dean John D. Hutson is interested in trying and promoting new initiatives that make lawyers better. Professor Sophie M. Sparrow brings teaching and assessment experience.
3. More information about the Webster Scholar program can be found at <http://www.students.piercelaw.edu/webster.pdf>, <http://www.piercelaw.edu/news/mediainfo/clippings/websterschol.htm>, or by contacting Program Director John Garvey at jgarvey@piercelaw.edu, 603-228-1541, or either of the authors.

LICENSURE IN MY IDEAL WORLD

by Susan M. Case, Ph.D.

The discussions about ways to improve licensure examinations are exciting discussions that we welcome within the testing unit at NCBE. While I have spent the past few years deeply entrenched in working with existing examinations, I appreciate being able to step back and think about my ideal

world of bar admissions. Of course, my ideal is affected by my training and experience in testing and measurement, and the concerns about fairness that necessarily follow. In my world, the examination would include the following attributes:

Content relevant to newly licensed practitioners. Every question would assess something that reasonable people agree a new lawyer should possess in terms of knowledge, skills, or judgments. This is not to say that every examinee would answer every question correctly, but that each question would be a reasonable one to include on the test. This attribute can be met by all examination formats, even multiple choice, if the questions are structured appropriately.

Broad content sampling. The content would be sampled broadly enough and be comparable enough across forms and test administrations that candidates would not be disadvantaged by receiving one form of the test as opposed to an alternate form of the test. Examinees should not feel that they were unlucky in the selection of content included on the test.

Accurate and valid grading. The scores would reflect the quality of the answers, and would not be affected by things that are not relevant (such as grader inconsistency).

Equated scores to ensure fairness across time. The scores would be equated over time so that the scores would maintain the same meaning regardless of the proficiency of the particular cohort of examinees and regardless of the relative difficulty of the exam form.



SUSAN M. CASE, PH.D., is the Director of Testing for the National Conference of Bar Examiners. From 1976 to 2001, she worked at the National Board of Medical Examiners conducting research and developing medical licensure and specialty board examination programs. These programs included various examination formats, including performance examinations simulating interactions in the real

world, computer-administered examinations, and written examinations. Her Ph.D. and M.S. degrees are in measurement and evaluation.

Reliable scores to ensure fairness across exam forms. The grades would be reliable enough that if a group of candidates were to be tested again, the rank-ordering of those candidates and their pass/fail outcomes would be very similar.

Anonymity of examinees to avoid bias. Answers would be graded without regard to the identity of the examinee. The grader would not be able to determine private information about the examinee such as name, law school, age, gender, or ethnicity.

Reasonable costs. The examination would be relatively inexpensive. By this, I don't mean that the exam should be cheap, but rather that a component that is more expensive than existing components would not be included without a rationale for incurring the additional expense. Research would have to indicate that the new, more expensive component added something to the measurement outcome that resulted in passing more people who should pass or failing more people who should fail.

MEDICAL LICENSURE

Many people both inside and outside the licensing world are interested in the standardized patient examination that was recently added to the licensing examination process for physicians. This component should be considered within the overall United States Medical Licensing Examination (USMLE) program, which requires that each examinee pass several separate hurdles: Step 1, a full-day multiple-choice test that most U.S. students take after their second year of medical school; Step 2, a full-day multiple-choice test that most U.S. students take near the beginning of their senior year of medical school; Step 2 Clinical Skills, a full day of interactions with standardized patients,¹ and Step 3, a one-and-a-half-day exam including a full day of multiple-choice questions and a half day of computerized simulations of patient interactions. Each component is offered in sites around the world, except the clinical skills examination, which, because of its significant cost and complexity, is offered in only six sites, requiring significant amounts of examinee travel.

Please note that each step in the examination sequence includes hundreds of multiple-choice questions that are used to evaluate the ability of examinees to apply their knowledge across a broad range of clinical situations; these questions span the content that is covered by the medical license, even though most physicians ultimately restrict their practice to something less than what is covered by their medical license. Because of the large number of questions, scores on the multiple-choice components are highly reliable; they are also equated across time so that scores maintain the same meaning regardless of the particular cohort

(Continued on page 30)

There are several things that are not listed above because they are not high on my list of desirable attributes. My ideal examination would not have to be a replica of real practice experiences. In order to provide standardization across examinees and anonymity of candidates, I am willing to give up this attribute. Verisimilitude is typically reduced when standardization and anonymity are more heavily emphasized. Multiple-choice exams, for example, are often criticized as being unrealistic; however, they successfully fulfill far more of the above attributes than any other type of exam.

Every testing format has strengths and weaknesses. Some formats that are optimal for educational purposes are less optimal for high-stakes licensure purposes. Table 1 on the next page comments on the extent to which five sample assessment formats (apprenticeship, standardized client examination, written performance test, essay examination, and multiple-choice examination) possess each attribute described above. They are listed in order of verisimilitude—apprenticeship is closer to real practice than is a standardized client examination, etc. Multiple-choice exams are the least like real practice, although questions that are framed within the context of real cases (such as those used on the MBE) require decision-making skills that are far closer to real life than the skills tested by the multiple-choice tests we all remember from school, basically recall of isolated and often picky facts.

The sidebar on this page and p. 30 that discusses the testing sequence for medical licensure is included in this article because critics of the existing bar examination often endorse parts of the medical licensure examination program as being superior to the licensure components in law. It is important to note that the medical licensure system has several hurdles,

TABLE 1

	Apprenticeship	Standardized Client	Written Performance Test	Essays	Multiple-Choice
Content Relevant to Entry Level Practice	Can be structured to be relevant.	Can be structured to be relevant.	Can be structured to be relevant.	Can be structured to be relevant.	Can be structured to be relevant.
Broad Content Sampling	Can be broad, but could also be quite limited in scope.	Usually is limited because of significant testing time per case.	Usually is limited because of significant testing time per case.	Usually is limited because of significant testing time per question.	Yes. Can sample 50 cases per hour.
Accurate and Valid Grading	Because each experience is unique, this is difficult to assess. Grading can be done appropriately, but can be biased and unfairly harsh or lenient.	Yes. Interactions and scoring keys are highly structured, so grading can be accurate and valid.	Yes, can be done appropriately.	Yes, can be done appropriately.	Yes
Equated Scores to Ensure Fairness Across Time	No. Because by its very nature, the assessment is unique to each examinee, there is no way to assure consistency across examinees.	No, because cases cannot be repeated over time, but can be scaled to objective tests given at the same time.	No, because cases cannot be repeated over time, but can be scaled to objective tests given at the same time.	No, because cases cannot be repeated over time, but can be scaled to objective tests given at the same time.	Yes
Reliable Scores to Ensure Fairness Across Exam Forms	Cannot ensure fairness across possible alternative apprenticeships.	Can be done, but requires days of testing time.	Can be done, but requires days of testing time.	Can be done, but requires days of testing time.	Yes
Anonymity of Applicant to Avoid Bias	No	No	Yes	Yes	Yes
Costs	High	High	Moderate	Moderate	Low

of examinees and the relative difficulty of the questions sampled. In addition, examinees are required to have graduated from an accredited medical school prior to licensure; many skills that are not assessed by the licensing examination sequence are assessed instead within the medical school experience.

After graduation from medical school, almost all graduates enter a residency program. After several years of residency, the graduates take specialty board examinations in order to become "board certified." While many medical specialties have oral examinations or other examination components, all have significant multiple-choice components to provide equated, reliable scores. It is also worth noting that the examinations are cumulative, in the sense that content from Step 1 may be covered again in Step 2, Step 3, and even in post-residency board examinations. For example, pharmacology is covered in Step 1 from a basic science perspective, and then is covered in increasingly sophisticated ways throughout the subsequent examination sequence.

Note: Keep in mind that for physicians, the test is national in scope, and entitles them to practice anywhere in the U.S., as opposed to lawyers who are admitted to one U.S. jurisdiction at a time.

1 Standardized patients are people trained to portray real patients. After eliciting information from the patient and performing a focused physical examination, the examinee records findings in a medical record. The examination includes interactions with a dozen standardized patients and requires a full day of testing.

some of which have multiple components. Every one of those steps includes one day or more of multiple-choice items.

There are strengths and weaknesses associated with every format that might be used for licensure purposes. Testing for licensure is intended to provide evidence that an individual possesses sufficient knowledge and skills necessary for entry-level professional practice. The multiple-choice format is the most common format for licensure exams in the professions, although professions may use more than one format in their examination sequence.

Non-multiple-choice formats, such as performance tests and even apprenticeship assessments, have much to offer by assessing skills of interest, such as the ability to communicate in writing or orally and the ability to ask relevant questions to identify the significant information in a legal problem. However, because of their limitations, such as low reliability, lack of anonymity, and lack of standardization, these formats should not be used in isolation. Combining scores across formats, and scaling scores on non-standardized components to the MBE helps to ensure that the scores are determined without bias and are comparable across time and across testing sites, which are important attributes in ensuring a fair licensing exam. 