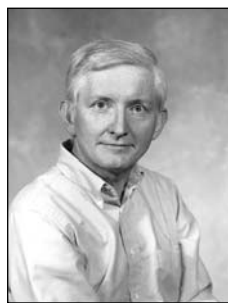# Clearing the Bar:

How to Set the Standard

Bar examining authorities, when deciding which candidates are minimally competent to practice law, must set a standard for such competence. Where to set the bar? What requisite level of knowledge, skill and judgment is sufficient to carry out the objective of protecting the public from ill-prepared or unskilled aspirants? In recent years, a number of U.S. jurisdictions have begun looking more critically at how they set their passing scores for bar examinees and most have solicited the help of testing professionals in the review process. In some of those jurisdictions, the standard-setting analyses have resulted in recommendations to the state supreme courts for changes in the passing scores. Not surprisingly, any recommendations for increases in the passing scores have been met with opposition from other players in the legal communities within those jurisdictions. Bar examiners, on the other hand, take seriously their charge to protect the public from incompetent would-be lawyers and note that failing applicants, in virtually every jurisdiction, have the option to retake the examination several times.

Because the area of standard setting is so important, and because most standard-setting processes involving bar examinations seem to set off controversy, the National Conference decided to create an "open forum" on the subject. We invited several individuals with experience in this area—testing professionals, law school professors or deans, and an admissions administrator—to write essays on standard setting. The authors were asked to write a one-thousand-word essay about any aspect of standard setting; they were not told who the other authors would be and they did not have an opportunity to review the other essays. As with other writings that appear in this publication, the views and opinions expressed, as well as the interpretations of data, are not those of the National Conference unless so stated. If you have reactions to any views presented here, please let us know. Write to Annie Walljasper, Editor, in care of the National Conference, or at awalljasper@ncbex.org.

Standard Setting for Licensure Examinations

by Michael T. Kane, Ph.D.

Standard setting continues to be a very contentious issue for high-stakes testing in general and for bar examinations in particular. The choice of a specific standard for admission to practice is necessarily judgmental and therefore debatable, and yet, any change in passing score has clear and immediate consequences for some candidates.

Bar examinations are designed to make decisions about whether individuals are prepared for practice. They require candidates to perform some tasks (e.g., answer

multiple-choice or essay questions), which assess the knowledge, skills, and judgment (KSJs) needed in practice. The justification for using scores on the examination to make decisions about admission to practice involves a chain of inference leading from a test score to conclusions about achievement on the KSJs, and thence to conclusions about readiness for practice.

Standard-Setting Methods

All standard setting relies on judgment, and the available standard-setting methods can be divided into two broad categories, *test-centered methods* and *examinee-centered methods*, based on the kind of judgment employed.[1] In test-centered methods, generally used for multiple-choice tests, panelists make judgments about test items. For example, in the Angoff procedure (the most commonly used test-centered method), panelists are asked to decide on the probability, called a *minimum pass level*, or MPL, that a hypothetical, minimally competent candidate would answer each item correctly. The MPLs are then summed across items to get the passing score on the test.

In the examinee-centered methods, used more often for essay and performance tests, panelists make judgments about whether or not examinees have the KSJs needed in practice, based on a sample of their performance. The passing score is then chosen so that it differentiates between the two groups as well as possible. The classic example of an examinee-centered method is the contrasting-groups method, in which panelists categorize examinees into two groups, those who have met the requirements and those who have not met the requirements. The passing score on the examination is then chosen to differentiate between these two groups.

Criteria for Standard Setting

Performance standards and passing scores have been accused of being arbitrary in two senses. First, the imposition of a specific cutpoint on a continuous scale is always somewhat artificial. Candidates with scores just above the passing score are not much better prepared than those with scores just below the passing score. This kind of arbitrariness, however, is inevitable whenever pass/fail decisions are based on a continuous score scale.

A second and potentially more serious concern is the articulation of the performance standard. In some cases it is fairly easy to specify an appropriate standard, given the goals and the context of the decision to be made. For example, a requirement that a passenger ship have at least as many life jackets as it has passengers and crew is clearly justified by a desire for safety, and the number of life jackets required is determined by the number of people on the ship. The standards for admission to a profession are not so easy to define. Given the wide range of contexts in which practice occurs, the passing score for a licensure examination is much harder to determine than the number of life jackets needed on a ship.

Current testing guidelines suggest that bar examiners should strive for a passing score that is "high enough to protect the public but not so high as to be unreasonably limiting"[2] and that "depend[s] on the knowledge and skills necessary for acceptable performance in the profession."[3] By linking the passing score to accepted standards of practice in the legal profession, the standard can be justified as having a "rational connection with the applicant's fitness or capacity to practice."[4]

To the extent that panelists have experience in evaluating performance in the profession, it will generally be in evaluating extended performances on professional tasks in practice situations. Therefore, if standard setting is to be based on the evaluation of candidate performance against accepted standards of practice, an examinee-centered approach seems preferable to a test-centered approach.

Implementation of Examinee-centered Standard Setting

Many issues in standard setting are still under debate, but it is possible to suggest how some generally accepted guidelines might apply to an examinee-centered standard-setting study for a bar examination. At least five aspects of an examinee-centered study have a direct impact on the passing score: (1) the choice of panelists; (2) the specification of the performance standard; (3) the training of panelists; (4) the choice of examinee performances to be evaluated; and (5) the implementation of data collection and analysis procedures. It is necessary that all of these issues be addressed in an appropriate way for the resulting passing score to be defensible.[5]

First, all standard-setting methods involve judgments and therefore all require qualified panelists. The panelists in an examinee-centered study must have enough technical expertise to evaluate candidate performances. Familiarity with the population of candidates for admission to legal practice and with the work of newly admitted lawyers should help to keep the standard realistic. In setting standards for a bar examination, good pools of potential panelists would include practicing lawyers who supervise newly admitted lawyers, law school faculty who are also involved in practice, and judges who regularly see the work of newly admitted lawyers. The number of panelists should be large enough to achieve an acceptably small standard error for the resulting passing score (probably 10 to 20 panelists). It is useful to employ two or more independent panels as a check on the reliability of the results.

Second, at the beginning of the standard-setting process, the panelists are expected to reach some level of agreement on a ***performance standard*** describing the level of competence required for entry-level practice, which is then used to identify an appropriate ***passing score***. The initial statement of the performance standard can be refined during the study, but it is important to start with a clear focus. The performance standards are likely to be most defensible if they are clearly linked to generally accepted standards of practice.

Third, the panelists should get thorough training on what they are expected to do. Training should continue until both the panelists and those conducting the study are satisfied that the panelists are comfortable with the performance standard, the assessment tasks, and the process to be used in translating the performance standard into a passing score. Previous experience in scoring examinations is useful but not sufficient.

Fourth, in an examinee-centered study, panelists evaluate a sample of performance for each examinee included in the study against the performance standard. The sample of examinees should be large enough (probably 50 or more) to provide stable estimates of the passing score. The sample of performance for each examinee should be extensive enough (e.g., three or four essays) to provide panelists with a good indication of whether or not the examinee has met the performance standard.

Fifth, standard-setting studies generally take from two to three days, with most of the first day devoted to orientation and training. The panelists then evaluate performances in batches, usually with discussion of their decisions after each batch. Panelists generally get to review and revise their decisions before the passing score is finalized, and are often provided with feedback on the conse-quences of their decisions, if such data are available. As an additional check on the design and implemen-tation of the standard-setting process, the panelists are asked to evaluate the standard-setting process and its outcome.

Concluding Remarks

Given the way bar examinations are used to make decisions, it is necessary to have a specific passing score. The aim is to provide adequate protection to the public, while not subjecting candidates to arbi-trary requirements, and therefore, the choice of a passing score on a bar examination is a matter of public policy. This policy is articulated in the per-formance standard and implemented through the passing score.

ENDNOTES:

1. R. M. Jaeger, *Certification of Student Competence*, in R. L. Linn (Ed.), EDUCATIONAL MEASUREMENT (3rd ed.) 485-514 (1989). New York: American Council on Education and Macmillan.

2. American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education, THE STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING 157 (1999).

3. *Id.* at 162.

4. K. M. Pyburn, *Legal Challenges to Licensing Examinations*, EDUCATIONAL MEASUREMENT: ISSUES AND PRACTICE, 9(4), 5-6 at 14 (1990).

5. *See* G. Cizek, STANDARD SETTING: CONCEPTS, METHODS, AND PERSPECTIVES, Lawrence Erlbaum (2001). This book of readings provides current views on a wide range of issues in standard setting.

MICHAEL T. KANE, PH.D., is the Director of Research for the National Conference of Bar Examiners. Before he joined the Conference, Kane was a faculty member in and chair of the Department of Kinesiology at the University of Wisconsin-Madison. At UW-Madison, Kane taught courses in measurement theory and practice and conducted research in educational meas-urement. He is a frequent speaker on psychometric topics, and he consulted regularly with the National Conference on testing issues before joining the organization. Earlier in his career, Kane was a senior research scientist at ACT, Inc., in Iowa City, Iowa; at ACT he provided psychometric consultation and support for ACT resident and contract programs, developed procedures for programs and services, and conducted research on psychometric topics. Kane received a Ph.D. in education and an M.S. in statistics, both from Stanford University, and a B.A. and M.A. in physics from Manhattan College and SUNY, Stony Brook, respectively.

RAISING THE BAR: LIMITING ENTRY TO THE LEGAL PROFESSION

by Deborah Jones Merritt

The 1990s witnessed the best of times and the worst of times for bar examinees. First the good news: Examinees who took the Multistate Bar Exam-ination (MBE) during the last decade of the twentieth centu-ry registered higher scores than those who preceded them. Nationwide, mean scaled scores on the MBE rose during the last sixteen years, from a July mean of 139.2 in 1984 to a July mean of 142.0 in 2000. In 1994, the July mean reached 145.2 and since then, July means have remained at 142.0 or above. Similar gains have occurred in February scores, although those scores are lower than July scores each year.[1]

Because scores from the multistate exam are equated over time, these numbers demonstrate an absolute rise in the quality of bar applicants—as measured by the exam itself. The MBE does not suf-fer from grade inflation. Examinees who took this

test during the last decade are more competent, as measured by the MBE, than those who took the exam during the preceding ten years.

More good news: At the same time that applicant quality has risen, aspiring attorneys have become more racially diverse. In 1984, less than nine percent of that year's law graduates were people of color. By 1999, that percentage had more than doubled, with students of color earning nineteen percent of all law degrees nationally.[2]

But now the bad news: Rather than welcoming these diverse, talented applicants to the profession, more than a dozen states have raised the scores needed to pass their bar exams in the past ten years.[3] These actions excluded from the profession individuals who were more competent—and more diverse[4]—than attorneys admitted during the previous decade. About one-third of exam takers from ABA-accredited law schools now fail the bar exam each year.[5]

Some states raised their passing scores by fiat, with little investigation or explanation.[6] Others attempted to set scores more scientifically, but used a flawed method designed by consultant Stephen Klein.[7] Lowell Hargens, Barbara Reskin, and I have explained the flaws in that method at greater length elsewhere,[8] but I review them briefly here.

The heart of Klein's method uses ratings from expert panelists to estimate the percentage of passing essays on a recent administration of the state's bar exam. The process then equates the percentage of passing essays with the percentage of passing bar exams in that year. Finally, Klein determines the passing score that would have generated that percentage of passing exams—and recommends that the

state adopt a score similar to that one.[9]

Klein's method stumbles on the assumption that the percentage of passing essays equals the percentage of passing bar exams. This assumption would be plausible if each essay on the bar exam measured the same competency, such as general problem-solving skill. One could then assume, as Klein does, that each essay score represented an independent measure of that single skill.

But essay questions on the bar exam measure much more than general problem-solving ability. Each question also measures competence in a specialized—and different—field of law. Exam takers who fail one essay do not necessarily fail other essays. Indeed, analysis of actual scores from a recent bar exam in Ohio demonstrated that essay scores show *low* inter-question correlations.[10] Because bar exam essays tap different competencies, and because most states allow applicants to fail several essays without flunking the bar,[11] the percentage of passing *essays* among each group of applicants has no necessary relationship to the percentage of passing *exams* completed by those applicants. Klein's standard-setting method founders on this fundamental flaw.

Klein's method might work if experts participating in his studies were asked to evaluate the entire package of essays written by an applicant—or if the process added both expert judgments about the number of essays that a minimally qualified applicant must pass and data about the distribution of passing and failing essays among applicants. In its current form, however, the method generates an arbitrary passing score.

Careful standard setting for the bar exam, incorporating the views of experts drawn from diverse

fields of the profession, is desirable. Indeed, properly designed exercises probably would demonstrate that the passing score currently is too high in many states. But standard-setting processes must pass expert scrutiny and make realistic assumptions about the bar exam.

Standards must also accord with common sense. Current standards, which fail one-third of examinees graduating from accredited law schools, don't pass the common-sense threshold. The American Bar Association applies rigorous rules to accredit law schools. Law students are a select group with premium qualifications. And legal education has become increasingly demanding in recent years. Professors who once administered a single exam at course's end are now assigning research papers, negotiation exercises, and midterms. An exam that flunks one out of every three students completing this rigorous three-year curriculum defies common sense.

Most regrettable of all, recent rises in bar passing scores have excluded increasingly diverse applicants from the profession. Because MBE scores are equated from year to year, and most states scale raw essay scores to the MBE, the rise in passing scores represents an absolute increase in the competency demanded of examinees—not a correction for some type of grade inflation. Examinees in many states must display a higher level of competence today than their peers demonstrated ten or twenty years ago.

Why does a group of applicants that is one-fifth nonwhite have to show a higher level of competence than their mostly white predecessors displayed twenty years ago? Advocates of higher passing scores have not answered this troubling question. Minority applicants don't need special standards to

pass the bar exam, but they shouldn't have to face a rising hurdle.

Bar exams protect the public from incompetent attorneys, but they also restrict competition and hamper the diversification of our profession. Recent increases in passing scores have served the latter, ignoble ends without sufficient evidence that they were necessary to achieve the former, noble one. It's time to reverse recent increases in the passing score and to search for standards that protect the public while treating today's bar applicants fairly.

ENDNOTES

1. All figures in this paragraph are drawn from *2000 Statistics*, 70 BAR EXAMINER 2 at 6, 21 (2001).

2. In 1984, law schools awarded 36,687 J.D. degrees; 3,169 or 8.6 percent went to minority graduates. In 2000, schools granted 38,157 J.D. degrees; 7,391 or 19.4 percent of the graduates were minorities. *J.D. Degrees 1984-2000 (Total/Women/Minorities), http://www.abanet.org/legaled/statistics/jd.html.*

3. States that have raised their passing scores include Arizona, Georgia, Illinois, Kansas, Maine, Missouri, Nebraska, North Carolina, Ohio, Oklahoma, Pennsylvania, Rhode Island, Texas, and Wisconsin. Other states have implemented more complex changes that may have had the effect of making bar passage more difficult. Only one state, New Mexico, plainly lowered its passing score during the 1990s. Two others, Mississippi and New Jersey, both raised and lowered their passing scores with the net effect unclear. Pennsylvania (which raised its score earlier in the decade) is considering a possible decline, although not below the 1990 level.

4. The precise racial composition of applicants in the "new fail zone," i.e., the group of those who fail today's bar exams but would have met previous passing scores, is unknown. But given the dramatic increase in minority applicants since 1984, it is almost certain that *every* stratum of bar applicants—including those falling in this zone—is more diverse today than it was in 1984. Thus, whether minority applicants make up five percent, ten percent, or 50 percent of the "new fail zone," that percentage almost certainly is higher than the percentage of minorities in that zone fifteen years ago.

5. *2000 Statistics*, 70 BAR EXAMINER 2 at 6, 10 (2001).

6. The Illinois Supreme Court, for example, raised the passing score on its bar exam without inviting public comment or revealing the method by which it had chosen the new passing score. *See, e.g.*, Chris Klein, *Illinois Deans' Dilemma—Is the Bar*

Exam (A) Sentry or (B) Aptitude Test?, NAT'L L.J., Dec. 30, 1996, at A16; Jackson Williams, ***Irreconcilable Principles: Law, Politics, and the Illinois Supreme Court***, 18 N. ILL. U. L. REV. 267, 318 (1998).

7. Klein first described his process in 1986 as one he had used to recommend a new passing score for the Puerto Rico bar exam. Stephen P. Klein, ***On Testing: Establishing Pass/Fail Standards***, 55 BAR EXAMINER 3 at 18 (1986). More recently, he has used the process to recommend passing scores for the Ohio, Florida, Minnesota, New York, and Pennsylvania bar exams. ***See, e.g.***, Stephen P. Klein, ***Panelist and Reader Judgments Regarding the Passing Score on the Florida Bar Exam*** (Aug. 12, 1999); Stephen P. Klein, ***Panelist and Reader Judgments Regarding the Passing Score on the Minnesota Bar Exam*** (Feb. 10, 1998); Stephen P. Klein, ***Independent Panelist and Reader Judgments Regarding the Passing Score on the Ohio Bar Exam*** (Jan. 5, 1996).

8. Deborah J. Merritt, Lowell L. Hargens, & Barbara F. Reskin, ***Raising the Bar: A Social Science Critique of Recent Increases to Passing Scores on the Bar Exam***, 69 CINCINNATI L. REV. 3, 929 (Spring 2001).

9. In addition to gathering information from expert panelists, Klein estimates the percentage of passing exams from post hoc judgments by actual exam graders. Those estimates, however, contain the same flaw as the one affecting estimates derived from the expert panels. Klein also compares the passing score in each state to that in other states. For a more detailed description of his process, see Merritt, Hargens, & Reskin, ***supra*** note 8, at 941-49.

10. Merritt, Hargens & Reskin, ***supra*** note 8, at 953-55.

11. Analyses of actual scores from one state, Ohio, show that the overwhelming majority of candidates fail some essay questions—and that minimally competent candidates fail multiple essay questions. ***See*** Merritt, Hargens & Reskin, ***supra*** note 8, at 959. States that require candidates to pass a minimum number of essays, moreover, tend to require candidates to pass just over half of the essays. Delaware, for example, which ranks with California as one of the toughest bar exams in the nation, allows passing candidates to fail five of its twelve essays. Rhode Island, which also maintains one of the nation's highest passing scores, likewise allows passing examinees to fail five out of twelve essays.

DEBORAH JONES MERRITT is the Director of the John Glenn Institute for Public Service and Public Policy at The Ohio State University, as well as the John Deaver Drinko/Baker & Hostetler Chair in Law and an adjunct professor of sociology. Merritt graduated from Harvard College ***summa cum laude*** in 1977 and Columbia Law School in 1980. While at Columbia, she was managing editor of the ***Columbia Law Review***, earned the distinction of James Kent Scholar, and won the Robert Noxon Toppan Prize.

After graduation, Merritt clerked for Judge (now Justice) Ruth Bader Ginsburg on the Court of Appeals for the District of Columbia Circuit and for Justice Sandra Day O'Connor on the Supreme Court of the United States. Merritt joined the law faculty at the University of Illinois in 1985 and moved to Ohio State in 1995. She has published widely on issues of federalism, equality, affirmative action, health and technology, legal education, tort reform, and law and social science. Much of her work has focused on public policy issues, and she has made numerous presentations to judges, legislators, and other policymakers. Merritt has also co-taught courses in Europe with both Justice Ginsburg and Justice O'Connor. She was appointed Interim Director of the John Glenn Institute in July 2000, and became Director in November 2000.

SETTING BAR EXAM PASSING SCORES AND STANDARDS

by Stephen P. Klein, Ph.D.

Before there was a Multistate Bar Exam (MBE), states asked their bar examinees several essay questions and then passed the candidates that had average scores of "70" or higher. Unfortunately, the difficulty of achieving a "70" could vary substantially from one administration of the exam to another due to differences in the average difficulty of the questions, differences in the leniency with which the answers were graded, or both.

Most states now address this problem by "scaling" their essay scores to the MBE's point system and basing the pass/fail decision on the sum of a candidate's MBE and essay scale scores. Most states also initially set the total scale score needed for passing so that it would correspond to about the same standard as that used before they began scaling.

Almost all states use a standard that they believe will screen out the candidates who are not yet "minimally competent" to practice. However, what constitutes minimum competency varies tremendously across states. To illustrate, Florida's and New

York's passing rates would drop about 30 percentage points if they suddenly adopted Delaware's standard, and California's rate would increase about 35 percentage points if it used Minnesota's or New Mexico's standard.

Several factors may affect a state's standard, including what it considers to be the level of proficiency that is necessary to begin practicing, whether the passing rate is publicly acceptable, and the expected effect on minority applicants. Candidate ability may matter too, and this varies greatly across states just as it does among law schools. Because of this variation, states with similar passing standards may have different passing rates and those with similar rates may have different passing standards. Hence, one cannot gauge a state's standard from its passing rate. State policies regarding section weighting, rereading, rounding, and score banking and transfer also may affect standards as can the relationship between a state's MBE and essay scores.[1]

A state's standard for passing an essay question may not coincide with its standard for passing the entire exam. For example, California instructs readers to assign a score of 65 to an answer that is a borderline pass, the readers are given detailed descriptions of the characteristics of the answers that fall at this and other score levels, and seasoned graders train and supervise the readers in how to use these standards. Nevertheless, the percentage of candidates having an average essay score of 65 or higher is about 10 percentage points less than the percent passing the whole exam. Thus, if California based its pass/fail decisions on this standard, as Merritt and her colleagues have suggested,[2] then California's bar passage rate would *drop* about 10 percentage points. California's use of the MBE and scaling prevent this from happening.

During the past decade, more than a dozen states raised the score they require for passing the bar examination. Ohio did this in part through the use of a method I recommended. All of the other jurisdictions that raised their passing scores relied on a variety of other procedures. Thus, contrary to Merritt's contention, the trend toward higher passing scores was not driven by any one standard-setting method. At least some bar examiners recommended raising cut scores because of concerns about possible score inflation on the MBE. Inflation could occur if students are receiving more practice in taking multiple choice exams in law school or if bar exam preparation courses are able to raise MBE scores without a corresponding improvement in the candidates' legal skills and knowledge. Such inflation would not be detected or corrected by the MBE's equating process. Thus, some states may have raised *passing scores* just to maintain the same *performance standards* as they used in the past.

The states that raised passing *scores* have not experienced the drops in passing *rates* that might otherwise be projected, perhaps because higher standards motivate students to be better prepared. For example, before Ohio raised its standard, its mean MBE score was well below the national average on every administration of the exam, but since it raised its standard, Ohio's mean has always been at or above the national average. There also is a critical distinction between the percent passing on a given administration and the percent eventually passing after one or more attempts. An increase in standards may affect the former rate without affecting the latter one. And it is the eventual rate that determines how many minority and non-minority candidates are licensed to practice.

I use a standard-setting method[3] that focuses on

the essay and performance test sections mainly because lawyers consider these sections to be the most valid parts of the exam. This method, which is similar to an approach suggested by Plake and Hambleton,[4] begins by discussing the factors standard setters should consider. For example, candidates are not expected to be experts in the subjects tested, they are taking the exam without resource materials or assistance, and they are under strict time limits and intense pressure.

The orientation of panelists to the standard-setting process also includes noting that no test is perfect. Some candidates who fail deserve to pass and vice versa. However, passing those who should fail is the more serious error because candidates who pass can practice without supervision as soon as they are licensed. Errors in the other direction are correctable by having candidates retake the exam, albeit several months later. Licensing boards must protect both the public and the candidates, but the public's needs come first.

The next step in my method involves the panelists' discussing what a passing answer to their assigned question should entail. They then evaluate a sample of answers that are drawn from the full range of quality in their state and assign a passing or failing status to each of these answers. Panelists do this independently to prevent a few of them from exerting undue influence over others. They do not know the scores that were assigned to these answers by the regular readers and they can pass as many answers as they wish. Some standard-setting methods inform panelists about how their initial judgments would affect the percent passing and then let them change their evaluations. I do not recommend doing this because it is tantamount to asking them to pick a passing rate.

The final steps in my method involve determining the reader-assigned score on each question that distinguishes between the answers the panelists did and did not pass, finding the average of these scores, and then identifying the total scale score that corresponds to this average.

The standard-setting methods I use are comparable to those suggested by other professionals in the field of testing, but are tailored to the particular needs of bar examiners. However, because there is no generally accepted way to set passing scores, there is no reason to rely on a single approach. That is why I encourage states to base decisions about passing scores on multiple sources of information (including the judgments of their regular readers and the standards used by other states) and to periodically review their standards and pass/fail rules.

ENDNOTES

1. R.L Linn & L.A. Shepard, *Item-by-Item Standard Setting: Misinterpretations of Judge's Intentions Due to Less Than Perfect Item Intercorrelations*. Paper presented at the Large Scale Assessment Conference, Colorado Springs, Colo. (June 1997).

2. D.J. Merritt, L.L. Hargens, & B.F. Reskin, *Raising the Bar: A Social Science Critique of Recent Increases to Passing Scores on the Bar Exam*, 69 CINCINNATI L. REV. 3, 929 (Spring 2001).

3. S. Klein, *Establishing Pass/Fail Standards*, 55 BAR EXAMINER 3, 18-30, (1986); S. Klein, *Setting Pass/Fail Standards on Constructed Response Licensing Tests*. Paper presented at the meetings of the National Council on Measurement in Education, Chicago, Ill. (April 1991).

4. B.S. Plake & R.K. Hambleton, *A Standard Setting Method Designed for Complex Performance Assessments: Categorical Assignments of Student Work*, EDUCATIONAL ASSESSMENT, 6, 197-215 (2000).

STEPHEN P. KLEIN, PH.D., has consulted for the National Conference of Bar Examiners and more than two dozen state boards of bar examiners. He has also consulted for licensing and certification boards in accounting, actuarial science, dentistry, medicine, psychology, teaching, and other professions. He has testified as an expert in state and federal courts and was appointed by a federal judge to serve as the court's technical advisor.

In addition to providing consulting services, Klein is a Senior Research Scientist at RAND. Earlier in his career, Klein served as a research psychologist with the Educational Testing Service in Princeton, New Jersey, and as an associate professor in residence at UCLA. He received his B.S. from Tufts University and his M.S. and Ph.D. from Purdue University.

Standard Setting: The Impact of Higher Standards on the Quality of Legal Education

by Steven C. Bahls

Setting the score necessary to pass a state's bar exam is one of the most difficult tasks facing bar examiners and state high courts. About five years ago, the Supreme Court of Ohio substantially raised the cut score for the Ohio Bar Examination. The result, as expected, was a significant reduction in the bar pass rate. This change in the cut score has affected the quality of legal education in Ohio, positively in some respects and negatively in others. The purpose of this essay is to identify the likely impact of changes in cut scores on the quality of legal education and to suggest ways to mitigate any negative impact.

Impact of Cut Scores on Admission and Retention Policies.

Law schools are naturally concerned about the performance of their students on the bar exam. Few, if any, ABA-accredited law schools will admit students who are not likely to pass the bar exam. The new Ohio cut score influenced Capital University Law School, and other law schools in Ohio, to reduce the size of entering classes. To do so, however, is not without problems. It is difficult to predict, based on undergraduate GPAs, LSAT scores and other available information, who will fail the bar exam. In years past, a significant number of Capital law graduates with lower undergraduate GPAs and LSAT scores successfully passed the bar exam and developed distinguished legal careers. Today, many of these potentially strong future lawyers are being denied admission to law school because of the need to raise the median GPAs and LSAT scores of incoming students to improve bar performance.

Many law schools responded to an increase in the cut score by increasing academic attrition. Most law schools find a high correlation between law school GPA and bar performance. Law faculty have been more stringent in grading so that those students who are unlikely to pass the bar exam are excluded from law school under a school's academic policies regarding maintaining a minimum GPA while in law school. Of course, these policies also tend to exclude "late bloomers," who might be good lawyers, from continuing in law school.

Impact of Cut Scores on Law School Programs of Instruction.

The decision to substantially increase the cut score in Ohio has strengthened the quality of legal education in some respects and weakened it in others. In states with high cut scores, law professors may be more demanding of their students and their students may therefore be better prepared; on the other hand, those law students will typically have more courses with multiple-choice exams and will take fewer clinical and perspective courses that may better prepare them for the realities of law practice.

At Capital University, law professors have responded to an increase in the cut score by demanding greater class preparation and by expecting better

performance on law school exams. Faculty at Capital have modified the curriculum to strengthen the legal writing program because of the belief that weak writing skills are a contributing factor to those who fail the exam. All of these changes have strengthened Capital's program of instruction.

Other changes have materially weakened our program. Law professors now give more multiple-choice exams. These exams do not test the critical skills of legal analysis and reasoning as well as essay exams do; judges and clients don't ask questions and then give the lawyer four or five answers to choose from. Instead, judges and clients ask open-ended questions, forcing lawyers to identify the salient facts, frame the issues and apply the law. Multiple-choice questions have inherent limitations in their ability to test these skills.

The increase in the cut score has resulted in increased enrollment in courses covered by the bar exam and soft enrollment in clinical, skills and perspective courses. Prior to the change in Ohio's cut score, students were more likely to take courses that would prepare them to be the best lawyers, including skills and perspective courses.

Finally, Capital, like others, has created a non-credit supplemental bar review course, focusing primarily on how to take exams. While the program is worthwhile in that it is designed to increase the bar passage rate, the funds could be much better invested in courses designed to train students in the skills and values that are critical to the legal profession.

Impact of Cut Scores on Diversity.

Prior to modifying cut scores, licensing authorities should consider the impact of the change on diversity within the profession. Though the Supreme Court

of Ohio has not traditionally maintained statistics on the bar performance by members of minority groups versus non-minority groups, individual law schools have. Virtually all Ohio law schools report that a far lower percentage of minority students pass the bar exam than majority students. Research is also needed on whether the bar exam has a disparate impact on individuals from lower socio-economic classes who have not had the benefit of the finest high schools and colleges. I do not suggest that we relax bar standards so all can pass. However, I do suggest that legal educators and bar examiners study whether otherwise qualified minority law students disproportionately fail the bar exam when higher cut scores are used.

Reflections on Appropriate Cut Scores.

Before formulating a new cut score, licensing authorities should engage in a dialogue with law school professors about the impact of the proposed change on the quality of legal education. They should discuss how bar exams should create incentives for rigorous education, without creating a set of circumstances that force law schools to "teach to the bar" at the expense of skills and values courses.

Each state should keep statistical data (by race, gender and age) about how different demographic groups perform on the bar exam. Law schools have made substantial strides in increasing diversity among future lawyers, but the bar exam remains a formidable barrier to the profession in realizing the diversity that is necessary for all segments of the public to have confidence in the legal profession. Licensing officials should seek to balance a demanding bar exam with the need to avoid creating artificial barriers to minority students who would otherwise have the credentials to be solid lawyers.

Though I am not a mathematician or psychometrician, techniques used by several states to adjust cut scores appear to be quite soft, even though the techniques are shrouded in statistical analysis. All studies used by state licensing officials to adjust cut scores, that I am familiar with, have omitted what I consider to be the most important test—whether practicing attorneys can pass the exam under actual exam conditions using a new cut score.

When adjusting cut scores, states should consider modifying exams to focus more on lawyer skills and less on memorizing legal doctrine. At about the same time the Supreme Court of Ohio adjusted the cut score, it took two positive steps to improve the exam. First, it adopted the Multistate Performance Test, a performance-based exam component that focuses less on memorization and more on lawyer skills. Second, it reduced the number of subjects that the exam covered, thereby eliminating the need to memorize more law in the less important areas.

Finally, when adjusting cut scores, states should take steps to make exams fairer. In Ohio, for example, the Supreme Court has worked with law school professors to review the topics within substantive areas that are "fair game" on the bar exam. This exercise has the potential of eliminating those "oddball" questions that test obscure areas of the law. Similarly, bar examiners should always pre-test essay questions and hire question drafters who have demonstrated credentials in drafting quality exam questions.

Legal education is a continuum from law school, to licensing, to professional development after law school. Through open dialogue, law faculties and licensing authorities should work together to ensure a tough, but fair, bar exam designed to admit only qualified individuals, but not to unduly exclude law students who could be tomorrow's leaders.

STEVEN C. BAHLS is the dean of Capital University Law School. Before joining Capital, Bahls was a faculty member at the University of Montana School of Law where he served as an associate dean from 1991-94. He practiced law with Frisch, Dudek & Slattery, Ltd., of Milwaukee, Wisconsin, before he began his career in academia.

Bahls has written a number of articles for law reviews, trade magazines and national magazines; he also has authored practice manuals, law reform proposals, statutory annotations and classroom materials. In 1995, Bahls was appointed Co-Chair of the Ohio Special Commission on the Education of Lawyers. He also serves as Director of the Ohio State Bar Foundation and as President of the American Agricultural Law Association. He is a member of the Ohio, Wisconsin, Montana and American Bar Associations, the Association of American Law Schools and the American Law Institute. Bahls earned a B.B.A., with highest distinction, from the University of Iowa and a J.D., with honors, from Northwestern University School of Law. He teaches in the area of business and entrepreneurship law.

EXAMINING PASSING EXAMINATION SCORES

by Margaret Fuller Corneille

Boards of bar examiners in every state or jurisdiction have at some point in their histories—in the distant past or very recently—considered the question of where to set the minimum passing score defining bar examination success. The question of how this process should be accomplished and how often it should occur has recently come under close scrutiny and heated discussion in a number of states across the country including Minnesota, Ohio, Georgia, Florida and Pennsylvania.

Standard 29 of the Code of Recommended Standards for Bar Examiners (adopted by the American Bar Association, the National Conference

of Bar Examiners, and the American Association of Law Schools)[1] states that a "thorough study should be periodically made of the bar examination results to determine its effectiveness, to discover defects and to suggest possible improvements in the bar examination system." The Code also states that each jurisdiction "should make the results of these studies available to other jurisdictions upon request." Although the Code has been in existence since 1959 (amended by the ABA House of Delegates as recently as 1987) and has a place of honor on the shelves of bar examiners across the country, the notion of periodically re-examining the passing score is, in many cases, not without controversy.

Some states, such as Virginia and Georgia, studied the issue and raised their passing scores with little or no comment or objection. In Ohio, on the other hand, the increase in the passing score, which was instituted gradually over a period of several years, has led to vocal objections from various segments of the bar and especially from the deans of Ohio's nine law schools. In other states, including Minnesota and Florida, where increases in the passing score have been proposed but not implemented, deans of the law schools have also opposed the increases.

The Minnesota process did not begin with preconceived notions about which direction, if any, the passing score might go but arose out of the Board's obligation to revisit periodically the important question of setting the pass/fail score. The Minnesota Rules for Admission to the Bar (MRAB) state that the purpose for which the Board is established is

"[t]o ensure that those who are admitted . . . have the necessary . . . competence to justify the trust and confidence that clients, the

public, the legal system, and the legal profession place in attorneys."[2]

The Board's authority permits it "to establish a minimum passing score" for the examination.[3] The validity and integrity of the examination depend upon the passing score reflecting a level of minimum competence for the practice of law in the state of Minnesota.

In the mid 1970s, when the Multistate Bar Examination was added as a component of the Minnesota bar exam, the Minnesota Board of Law Examiners set its passing score at 260 (out of 400 possible points). The Board did not examine the issue again until 1998 when it retained Stephen P. Klein, Ph.D., to design two separate but related studies to consider the issue. In choosing Klein, the Board sought the advice of a testing professional who had conducted extensive research and had been published frequently over the previous 20 years on the topic of bar examinations and other forms of standardized testing.

Dr. Klein's first study was based upon asking the eight attorneys who acted as "lead graders" for the July 1997 Minnesota bar exam to re-read the essay answers from that exam, and to identify, on a 7-point raw grading scale, the score that represented a minimum passing score on the question.[4] The scores identified by each of the graders were then averaged and converted from the 7-point scale to a 200-point scale.[5] (In addition to the 200-point essay portion of the exam, Minnesota administers the Multistate Bar Exam, a 200-point multiple-choice test. Each portion of the test is weighted equally, thus totaling a 400-point scale.) The lead graders' score averages were then extrapolated to the 400-point total exam score range.

Using this method, Klein concluded that the graders' responses showed that the minimum passing score should be set at 272, rather than the 260 currently used in Minnesota. The 12-point difference between the existing Minnesota passing score and the graders' findings was significant and indicated a need for further study.

The second and more comprehensive study designed by Klein involved extending a public invitation to members of the bar in Minnesota to participate in a review of the bar exam passing score. Notice was published in the legal newspapers, and later, when the publication did not result in a sufficient number of volunteers, scores of letters were sent to attorneys asking them to consider participating in the study. The solicitation letters were sent to law firms and sole practitioners around the state, public law offices, law deans, and judges in an attempt to create a panel of reviewers who would represent a diverse cross-section of legal practice in Minnesota.

Eventually, a panel of 36 attorneys was assembled and included 26 practitioners, 6 law professors, and 4 judges. The panel was geographically diverse and included 7 members of minority groups; 27 of the panelists were men and 9 were women.

To conduct their work, the panelists met for a half-day Saturday session at the Minnesota Judicial Center. The process began by Klein's asking the participants to grade 35 bar exam essays that had been written by examinees who took the July 1998 Minnesota Bar Exam. Klein had pre-selected a complete cross section of answers based upon the original graders' scoring decisions.

Participants were invited to discuss the question and the range of answers with their fellow graders, but were asked not to attempt to reach consensus on grading standards. Participants were asked to read and assign a grade to each of the 35 essays using a 1 to 4 rating scale. The raw scores were described as follows: a score of 4 was "clearly passing"; a score of 3 was "minimally passing"; a score of 2 was "minimally failing"; and a score of 1 was "clearly failing."

Panelists were instructed to rate the papers based on the type of legal reasoning and writing that they would expect from any newly admitted attorney in order to meet a standard of minimum competency.[6]

The participant's ratings on the 1 to 4 scale were then converted to the 400-point scale for comparison purposes. Klein concluded that the study showed that the panelist graders, using the 4-point rating described above, would have established the Minnesota Bar Exam passing score at a level of 270.5—10.5 points above the current passing score of 260.

Taken together, these two studies suggested to the Board that some examinees were passing the Minnesota Bar Exam on the basis of essay answers that did not, in the opinion of a cross section of attorneys, judges and law professors, show a minimum level of legal competency. Accordingly, the Board recommended to the Supreme Court that Minnesota increase the minimum passing score from 260 to 270 over the following five years.

After receiving this recommendation, the Minnesota Supreme Court asked that the Board seek additional input from the legal community. The majority of those who responded, including the deans of the three Minnesota law schools as well as various representatives of the bar and legal organizations, objected to the proposed increase saying that

no studies had been done which showed that there was a lack of competence among newly admitted members of the bar. Many also raised concerns that the proposed increase would adversely impact minority law graduates. Others questioned the methodology and design of the process that had been used. After reviewing these comments, the Board concluded that the issue required additional study and that such study would be conducted before any recommendation for changing the score was implemented. The study of this issue continues today in consultation with and with the participation of the bench, bar and legal education communities.

ENDNOTES

1. Published in the ***Comprehensive Guide to Bar Admission Requirements.***
2. Rule 1, Minnesota Rules for Admission to the Bar.
3. Rule 3B, Minnesota Rules for Admission to the Bar.
4. Identifying a minimum passing score on each question is not normally included in the process of grading essay questions. Instead, graders score papers in rank order—from one to seven—with the worst papers designated as 1's, the best papers designated as 7's, and the remaining papers falling, in rank order, between those two points.
5. The essay portion of the exam is expressed on a 200-point scale in the same manner as the MBE portion, thus totaling 400 points.
6. The panelists were specifically instructed not to consider the performance they personally would expect from a newly hired attorney who worked for their firm or organization. Rather, they were asked to grade papers on how they thought any new attorney should be able to perform in order to meet a standard of minimum competency.

MARGARET FULLER CORNEILLE is Director of the Minnesota Board of Law Examiners, the Minnesota Board of Continuing Legal Education, and the Minnesota Board of Legal Certification. Prior to beginning work for the Minnesota Supreme Court in 1987, she directed Western Reserve Legal Services in Akron, Ohio, a civil legal services program providing legal services in northern Ohio. Corneille holds a J.D. from the University of Akron School of Law and a B.A., ***magna cum laude***, from Ohio University. She is co-chair of the Bar Admissions Committee of the American Bar Association's Section of Legal Education and Admissions to the Bar, and a member of the National Conference of Bar Examiners' Test Accommodations Committee.