

VOCABULARY AND BASICS

New York/MBE Scoring: New York compiles its results on a 1000 point scale using the MBE's scaling methods and equating the essay score distribution to the pool's MBE performance (see below for Dr. Kane's explanation of equating). New York scores can be converted to MBE 200 point scale equivalents by dividing by 5. Thus New York's current passing score (known as the "cut score") of 660 is equivalent to a 132 on the MBE scale, and the new cut score of 675 now being phased in will match an MBE scale 135. In accumulating points, scores on the five essays on Day One (no longer six) and the Multistate Performance Test (MPT) are combined and given 50% weight; the MBE on Day Two receives a 40% weight; the balance of the score comes from the NY-specific multiple choice (NYMC) at 10%. After raw scoring and any internal adjustments, the score distributions of the essay/MBT and of the NYMC are both "equated" to the current MBE mean and distribution before the final points are assigned.

Millman Report: a comprehensive evaluation, chaired by Jason Millman, of many aspects of the NYS Bar Examination, commissioned by the NY Court of Appeals and completed in May 1993. A panel evaluation of candidate essays from the July 1992 examination was one part of the evaluation, and the Millman Report concluded that the existing 660 (132) cut score was an appropriate passing standard. Content analysis was limited to NY portion of exam. Millman concluded that the essays had high content validity, but recommended some changes in the topics to be covered. Millman found that candidates' relative rankings on the MBE, essays and NYMC were highly correlated ($r = .88$ to $.93$); there is a wide body of research finding that rank order of candidates on MBE and essays in various jurisdictions is highly correlated.

Klein NY Study: a study of the July 1999 essay portion of the NY exam conducted by Stephen Klein for the NY Board of Bar Examiners, a report on which was included in the Board's 2002 Report on the Passing Standard. Klein asked the 36 attorneys who had actually read and scored the July 1999 essays (the "Reader Group") to use their experience with the pool of essays on their exam question to specify a minimum score that would "pass" that essay question. Klein also enlisted a group of 30 attorneys, judges and educators who were divided into panels for each essay question (the "Panelists"). Panelists looked at selections of 40 essays for their essay question and were asked to rate these as passing or failing on "minimal or basic competence." The essays they reviewed were drawn 2 each from 20 score "strata." Thus, only a handful came from the ranges the Readers considered marginal. From this process Klein concluded that the NY passing standard was too low because only about 55% of candidates should have passed the July 1999 exam rather than the 68% which did pass at 660 and that 690-93 (138-39) was an appropriate passing standard. Like Millman, Klein did not evaluate performance of NY candidates on the MBE or NYMC, but based his recommendation on extrapolating from essay performance.

Board Reports: The New York Board of Bar Examiners issued an initial "Report and Recommendation ... Regarding the Passing Standard" in March 2002 (Initial Report). Its Final Report on the topic was issued in December 2004 and is available at www.nybarexam.org. It contains discussion of the numerous comments on the initial Report and Recommendation, including responses by Dr. Klein to some of the criticisms of his methodology.

ABCNY Report: The Report of the Committee on Legal Education and Admission to the Bar of the Association of the Bar of the City of New York addressing the Board's Initial Report is available in the ABCNY's Record, Winter 2003, and at the abcny.org website under the Committee's heading. I am the current chair of this Committee, but my comments today are personal in nature and should not be attributed to the Committee or the Association.

Content Validity: Does the test cover topics and skills which are considered relevant to subsequent performance? Is the attention to each topic weighed reasonably, and do the questions test relative mastery of such subjects and skills in a way that meaningfully ranks candidates in terms of such mastery? Assessing content validity can involve evaluation of the test content by practitioners or educators. The Millman Report included such an assessment. Although opposing any increase in the cut-score, the Committee on Legal Education of the New York State Bar Association recently acknowledged that, despite "admitted shortcomings ... the bar examination has been a good proxy for the full-range of lawyer competences necessary to the practice of law" (NYSBA Submission in Opposition at 10). Content validity sometimes is tested by correlation with other sorting measures, but in the bar examination context this has been limited to prior performance (LSAT, law school grades) or contemporaneous performance (candidate rankings on the MBE are highly correlated with rankings on the essays), rather than subsequent performance in practice. The NY Board concluded that "a post-exam assessment of marginally passing candidates is not practical" and would be subjective and unreliable (Final Report 22), stating that bar exams assess competence at the time of examination, and "are not intended to predict the future performance of an attorney in actual practice" (Final Report 28).

Reliability Concerns (Consistency of Measurement): If the same test is administered again to a similar pool of candidates (theoretically, the same candidates) will the results (especially in the ranking of the candidates) be similar (error of measurement)? What would be the effect if different readers had read a particular candidate's essay answers? If subparts of a test are meant to test a common skill, how similar is candidate performance across the subparts (construct validity)? If a different version of the test had been administered (or is subsequently administered to a similar pool), will it produce similar results (reliability)? Since different versions of tests will always vary somewhat in difficulty, is there a mechanism for "equating" the tests from year to year?

Performance Standard: "The performance standard is intended to provide an answer to the question of how much is enough (i.e., what level of performance evidencing the skill or knowledge being assessed by the test is to be considered adequate for a particular purpose); such questions are often hard to answer." (M. KANE, THE BAR EXAMINER, 8/02 at 16)

Board's Statement on Cut-Score: "The passing standard must accurately reflect the minimum level of competent performance required for admission into the profession, in order to ensure that valid pass/fail decisions are made and the public protection function of the licensure process is thereby served" (Final Report 48).

Board's Statements of Problem: (1) Periodic standard-setting review is itself appropriate (Final Report 13-14); (2) concern with possible "MBE drift" as described by Klein and that such drift and equating method for essays might mask decline in performance in NY portion of test (Initial Report 15-16); (3) "There is already evidence sufficient to raise a concern about the

competence of newly admitted lawyers and to imply that additional public protection may be warranted” (Final Report 15).

Klein Statement of “Drift” concern: “At the time New York was considering adopting scaling, its board members were concerned that there might be a general increase in MBE score without a corresponding increase in the applicants’ essay writing skills ... I also said this problem could be addressed by periodically reviewing the passing score” (Klein May 2003 response, in Board Final Report).

Michael Kane Explains “Equating” of Pool Essay Performance to Pool MBE Scores:

“Because of its structure, the MBE can be statistically equated from one administration to another, thus promoting fairness, by subjecting all candidates to the same standards. Essay questions cannot be reused in high-stakes testing programs. So, each new form of an essay test typically consists of new questions, and perhaps new graders, so it is essentially impossible to avoid variability in the difficulty of essay examinations. To solve this problem, the MBE is carefully equated and then the essay scores are scaled to the MBE scores. This is not a perfect system. As you point out, it could happen that performance on the essay part of the test would get better or worse compared to the MBE, and this would not be visible after the rescaling. ... Individual candidates may do better on the essays (e.g., because they write well) or worse on the essays (e.g., because they don't write well), but it is much less likely that a whole cohort of candidates will do a lot better or worse on the essays and not show a corresponding change on the MBE. But if we did not control for the difficulty of the essay test by scaling to the MBE, the essay scores could fluctuate substantially (e.g., because a particular question turned out to be more difficult or easier than expected or because a new grader was unusually severe or generous). The system under which the MBE is statistically equated from one administration to the next and the essays are scaled to the MBE follows standard psychometric practice. It ... is designed to promote consistency in standards and therefore fairness.”

(taken from email kindly sent to me by Dr. Kane).

ANTICIPATED CONSEQUENCES

(1) Past History-Selected Statistics

(a)	All-takers passage rates	<u>660</u> (existing)	<u>675</u> (planned)
	July 1999	68	63
	July 2000	68	59
	July 2001	72	64

(b) 2002-2004 July exams: 83% of first-time takers from ABA schools passed; at 675 (assuming no other changes), pass rate would have been 78%.

(c) the Board's 675 is substantially below the 690-95 recommended by Dr. Klein. At 660 some 68% of candidates passed the July 1999 exam; Dr. Klein's recommended cut-off would have reduced this to 54-56%.

(2) National Averages Absent performance changes, the 675 cut score will bring New York's first time and total pass rates in line with national averages for candidates from ABA-approved law schools. Thus, using 675 on the July 2004 exam 69.6% of all ABA-school takers and 77.4% of ABA-school first-time takers would have passed.

(3) A Special Consideration New York's existing and anticipated pass rates look much tougher than the ABA-school statistics because approximately 25% of NY candidates are foreign-educated candidates qualified to sit for the exam under a special rule. New York tests over 95% of such applicants who sit for bar exams in the US, and this 25% proportion represents a doubling over less than a decade, from 12% in 1996. Such NY candidates achieve markedly lower MBE scores (131.2 in 7/01 versus 142.6 for ABA candidates) and markedly lower NY pass rates (in July 2004, 38.6% of foreign-educated candidates passed while 75.9% of ABA-school candidates passed).

(4) The Board's Expectations Notwithstanding that MBE scores nationally have remained relatively stable over more than two decades, the NY Board appears to be convinced that with greater diligence in examination preparation (and perhaps an increase in re-taking of the exam), over the longer term New York pass rates will return to prior historical levels, at least for ABA-school graduates. "We expect that most, if not all, of the same candidates who would pass the bar exam under the current standard will pass under the standard we have now adopted." (Final Report 33) "The Board proposes to partner with the law schools ... to improve the performance of their students who may be at risk of failing the bar exam" (49). "Specifically the Board anticipates that all candidates will promptly adjust ... By additional study and preparation, they will be even better prepared to pass ... when admitted, these new lawyers will enhance the profession with their ... competence" (44). The Board clearly believes that "additional study helps the candidates who marginally fail on their initial attempt at the bar exam" not merely to pass a subsequent exam but "to become better prepared to enter the practice of law" (34).