Pace Law Review

Volume 23
Issue 1 *Winter 2002*

Article 4

January 2002

A Critical Inquiry into the Traditional Uses of Law School Evaluation

Steven Friedland

A Critical Inquiry Into the Traditional Uses of Law School Evaluation

Steven Friedland*

TABLE OF CONTENTS

* I would like to acknowledge the useful assistance of Olympia Duhart, Sarah Yoho, Andrea Zigelsky, Wayne Clark and Jeffrey Berman with drafts of this article.

"In a very real sense, tests have invented all of us."[1] — F. Allan Hanson

I. INTRODUCTION

In an age of perpetual testing, it is at first glance surprising that law student evaluation has been widely overlooked. After all, testing has been a steadfast fixture of legal education since the late 1700s when several visionary individuals created the Litchfield Law School.[2] Yet, in the past half-century at least, comparatively few institutional resources have been devoted to the evaluation process,[3] and both institutions and individual instructors generally hold the evaluation process with a similar lack of regard. This article argues that the institutional and

1. F. ALLAN HANSON, TESTING TESTING: SOCIAL CONSEQUENCES OF THE EXAMINED LIFE 3, 4 (1993) ("The social person in contemporary society is not so much described or evaluated by tests as constructed by them. In addition to constructing social persons, tests . . . function to control and dominate them.").

2. Steve Sheppard, *An Informal History of How Law Schools Evaluate Students, With a Predictable Emphasis on Law School Final Exams*, 65 UMKC L. REV. 657, 665 (1997). Dean Christopher Columbus Langdell of Harvard Law School instituted exclusionary examinations in the 1870s—preventing students from graduating if they did not pass—perhaps to counter the low rates of success on graduation exams imposed on students. *Id.* at 671–72.

3. *See generally* Linda R. Crane, *Grading Law School Examinations: Making a Case for Objective Exams to Cure What Ails "Objectified" Exams*, 34 NEW ENG. L. REV. 785 (2000) (advocating the use of objective examinations in law school to increase the validity and reliability of the examinations as compared with the traditional law school essay exam); Daniel Gordon, *Does Law Teaching Have Meaning? Teaching Effectiveness,Gauging Alumni Competence, and The MacCrate Report*, 25 FORDHAM URB. L.J. 43 (1997) (discussing the lack of useful feedback from professors); Daniel Keating, *Ten Myths About Law School Grading*, 76 WASH. U. L.Q. 171, 171 (1998) (describing student and faculty-held myths about law school grading); Philip Kissam, *Law School Examinations*, 42 VAND. L. REV. 433 (1989) (analyzing the adverse effects of bluebook testing and suggesting changes).

faculty disinterest in law student evaluation is attributable in part to a triumph of formalism over functionalism. Like a monument in a town square that has long since lost its meaning, the evaluation process has become more highly valued for its perpetuation of rank and hierarchy than for its accuracy of measurement or its pedagogical attributes.

The aged monument status of law student evaluation can be gleaned from several observations. Traditionally, student evaluation—meaning the mechanisms, devices, or methodologies for judging a student's performance or potential[4]—has been embodied in a single final examination at the conclusion of a course.[5] By virtue of its location alone, the law school evaluation process has remained an appendage to legal education, often considered an "afterthought."[6]

In addition to its locus, the lack of institutional respect for the evaluation process has proven to be an even weightier burden. The unstated assumption is that evaluation is a necessary evil[7] for faculty members, an essentially unproductive element of the law school program, particularly in comparison to the scholarship or teaching enterprises.[8] Almost the entire evaluation process, save some administrative guidelines, rests on the

4. Evaluation and assessment often are used interchangeably when discussing testing mechanisms, but the two terms can differ in meaning. Assessment, especially for the purposes of this paper, is intended to center on observation and reflection of some performance or conduct, including thinking. Evaluation, on the other hand, is intended to include more than just observation of fact. Evaluation also has a normative aspect involving comparisons and judgmental distinctions. Thus, in the context of legal education, evaluation generally is the appropriate term for any graded mechanism, from in-class and take-home examinations, midterms, papers, and quizzes, to oral tests or other written assessments.

5. Crane, *supra* note 3, at 786 ("During the typical law school examination, students are asked to demonstrate their ability to recognize complex bundles of information and to perform well on a single test that is worth 100% of their grade . . . . ").

6. Mary E. Keyes and Michael J. Whincop, *The Moot Reconceived: Some Theory and Evidence on Legal Skills*, 8 LEGAL EDUC. REV. 1, 2 (1991); *see also* JOHN HEYWOOD, ASSESSMENT IN HIGHER EDUCATION, at vii (1977) (stating that assessment is the "grand afterthought of the educational process").

7. The status of law school evaluation is not consonant with the approach in other educational fora. In secondary, college, and many other graduate education curricula, evaluation usually occurs during, and even throughout, the substantive educational process, as well as at its conclusion.

8. It is almost axiomatic that the student evaluation process is not accorded the same institutional status as other components of legal education—or given a status equivalent to evaluation in other educational fora.

shoulders of individual instructors, who act without coordination, institutional oversight,[9] or validation safeguards.[10] The instructors are accorded this responsibility usually without formal or informal evaluation training,[11] input from veteran faculty members, or other institutional guidance, unlike the oversight of and support for a professor's teaching, scholarship and even committee work.[12] Instead, the primary institutional message is a simple one: get the grades in on time.

A third observation involves the close ties between the evaluation process and tradition.[13] Many teachers, recalling and repeating their own school experiences with evaluation, view examinations merely as negative motivation devices for students as well as a means of measuring thinking ability.[14] The validity, reliability, and rationale for the form and substance of evaluations are rarely articulated; safeguards assuring the same are just as unlikely.

---

9. Professors are often completely unaware of other professors' evaluation methods and examinations. Of course, it might be argued that the tendency is for professors to focus solely on their own courses and to have little, if any, knowledge about the direction and management of other courses, from the pedagogy of those courses to their evaluation processes.

10. The instructor is responsible for deciding on the form and number of test items, creating the content of the test, grading it, and then ranking the students.

11. *See* Crane, *supra* note 3, at 804. Most faculty members lack an understanding of psychometrics and the rudiments of creating a valid and reliable examination. The faculty is also entirely unaware of the evaluation process in other classes, even those classes in the same subject area. Professors do not know what evaluations are created and delivered, how those evaluations are graded, and how feedback is delivered.

12. The Association of American Law Schools (A.A.L.S.) offers a New Law Teacher's Workshop every year, which entails two full days of introduction to teaching and writing. There is generally one session, from forty-five minutes to an hour, on the subject of law school evaluation. Otherwise, there is little in the way of formal analysis of, or preparation for, the evaluation process. Instead, the instructor is welcomed into the world of evaluation and assessment usually at the end of the first semester of teaching, with little institutional guidance or support.

13. The bias against the evaluation process is historic as well as pragmatic. The feeling that "if it isn't broke, don't fix it," definitely applies to this component of the legal education process.

14. As a commentator noted: "In one study most academics surveyed saw assessment only in terms of it providing an incentive to make students work, and to enable their intellectual abilities to be measured." Jeffrey W. Barnes, *The Functions of Assessment: A Re-examination*, 2 LEGAL EDUC. REV. 177, 179 (1991) (citing David Watkins & Barry Morstain, *The Educational Orientations of Lecturers and their Students: A Case Study of an Australian University*, 24 AUST. J. EDUC. 155, 160 (1980)).

The consequences of this aged monument status have been dramatic. Law schools and instructors have undervalued assessment as a teaching tool[15] and overvalued evaluation as an accurate, objective measuring device.[16] Without institutional oversight, these problems have festered and multiplied.[17] The lack of clarity associated with the grading process, for example, has only contributed to its mystique, prompting one associate dean to declare, "[t]here is probably no subject more misunderstood and more clouded in myths than law school grading."[18]

Despite these problems, the aged monument status is valued highly by certain constituencies. Evaluation matters considerably to law students,[19] who see good grades as a primary path to employment opportunities.[20] When students come to understand the extent of the power wielded by evaluations,[21] particularly in comparison to the classroom experience,[22] the

15. Evaluation would matter greatly to the teaching enterprise if it were used to improve the efficacy of the learning process through feedback, information, and extra practice.

16. Exams are separated from mainstream pedagogy in most courses, yet still affect all that occurs in it. Students are focused on the examination, as they fully understand that the exam will dictate who will be considered a success in the course, and who will not. Even though the exam is merely summative of the learning process, it serves to motivate and guide students throughout the course.

17. For example, with the exception of research and writing classes, evaluation as a feedback mechanism— to help students improve their skills—rarely occurs during the substantive component of a course.

18. Keating, *supra* note 3, at 171.

19. Just ask students about the impact of examinations. Examinations have become almost mystical, magical events in the popular culture, glorified in movies such as the "Paper Chase" as the apex of the first year of school and of later years as well.

20. The need to perform well, especially in the first year of school, heightens the importance of the evaluation process. Students are cognizant of the fact that the grades are often the sole arbiters of future benefits, such as law review eligibility, clerkships, and permanent employment.

21. On a more localized basis, law school evaluation offers a reflection of a law school's culture, both on the narrow scale of how the law school approaches its honor code, and the broader scale of how the faculty views its students. Grades on an examination implicitly reflect the perceptions of teachers about students and vice versa. Also, the examination process provides a measure of a teacher's success in teaching insofar as it is a representation of how well the students are learning.

22. To many students, the classroom experience appears to have a more tenuous nexus to future success after grades are returned from the first semester. This lack of connection appears even more pronounced if grades do not appear to correlate to classroom performance.

behavioral control exerted by examinations expands even more. Students see that success often begins and ends with grades.[23]

The influence of grades has turned examinations in the eyes of law students into transformative devices, evolving from mere instruments of authority, whose primary function is ranking and ordering students, to a pervasive power defining the entire legal education experience.[24] In effect, evaluations construct much of a student's law school education. Evaluations create the successful student, the one invited for further honor and attention through law review, the one who will work for professors and judges, and the one who will obtain the most competitive jobs.[25] Examinations also loom large in creating self-perceptions about abilities, interests, and potential for success.[26]

Until law schools recognize the pervasive influence and power of law school evaluation on an institutional level, they will not elevate the functionality of evaluation over the formalism in which it is embedded. This article advocates[27] rethinking the use of such an important tool,[28] conceptualized and expan-

---

23. Students learn to "take the professor, not the course," and moderate their preparation accordingly. This means that a student will prioritize the course in the professor's image and try to anticipate the nature and type of questions on the exam.

24. *See* HAROLD BRODKEY, STORIES IN AN ALMOST CLASSICAL MODE 221 (1988):

> But I did well in school and seemed to be peculiarly able to learn what the teacher said . . . and there was the idiotic testimony of those peculiar witnesses, IQ tests: those scores invented me.
>
> Those scores were a decisive piece of destiny in that they affected the way people treated you and regarded you; they determined your authority; and if you spoke oddly, they argued in favor of your sanity.

Id.

25. Evaluations create reputations about general intellect, and more than that, how well-qualified or well-suited students are for particular areas of law.

26. In a burgeoning new field, some law professors are devoting time and attention to analyzing the negative influences of legal education on law students' mental health and overall well-being. This humanism movement is now represented by a section in the American Association of Law Schools who participated in a May 2001 conference on Therapeutic Justice at the University of Cincinnati Law School.

27. This article incorporates what the British philosopher Gilbert Ryle and the anthropologist Clifford Geertz called a "thick description" of history, economics, and sociology. CLIFFORD GEERTZ, THE INTERPRETATION OF CULTURES 6 (1973).

28. Prior to implementing changes, however, it is beneficial to first uncover the causes of the current situation. The issue of causality is embodied in a question: Why, despite the significant and far-reaching effects of examinations on stu-

sively reframed[29] as both a rigorous measuring instrument and a feedback tool.[30]

The paper is divided into four sections. A background section that recites the current status of evaluation principles and practices, and describes evaluation as a tool of control and social construction follows this introduction. In the third section, the article describes the problems associated with the current status of law school evaluation, in which exams are overvalued for their measuring capabilities and undervalued for their feedback qualities. The fourth section proposes a true dualist approach to the evaluation function, enhancing evaluations as measuring devices and expanding evaluations to serve as pedagogical tools.

II. EVALUATION PRINCIPLES AND THEIR APPLICATION TO LAW SCHOOL

"We are entering the age of the infinite examination and of compulsory objectification." —Michel Foucault[31]

A. *Defining Principles*

Evaluation and assessment are often used synonymously, but are not necessarily identical. Evaluation means "to judge [the] value, quality, importance, extent, or condition" of something.[32] It has a normative element—the judging and valuing of something. An evaluation may include a ranking, hierarchy, or appraisal. An assessment, on the other hand, can be an evalua-

---

dents, employers, teachers, and others, is evaluation routinely ignored, cabined from both scholarly opinion and practical observation alike? The answer to this question comprises the first part of this article, which argues that history, tradition, lack of training, and job-related disincentives contribute to the lack of status of evaluation in the constellation of legal education.

29. A famous "Saturday Night Live" skit involved a mock advertisement promoting a product as both a floor wax and a pie filling. While evaluations have the potential to be just as versatile, they generally have been viewed as having a singular role—that of a measuring instrument. This configuration can be altered to include the evaluation as a teaching and learning tool.

30. As some commentators have noted, "the sole valid purpose of any grading system should be to encourage maximum educational achievement and learning on the part of students." Stacy L. Brustin & David F. Chavkin, *Testing the Grades: Evaluating Grading Models In Clinical Legal Education*, 3 CLINICAL L. REV. 299, 306 (1997).

31. MICHEL FOUCAULT, DISCIPLINE AND PUNISH 189 (Alan Sheridan trans.) (1977).

32. MICROSOFT ENCARTA COLLEGE DICTIONARY 494 (2001).

tion, but it also can be defined more neutrally. Assessment can also have a descriptive meaning, such as reflection without the normative judgment, when a person simply communicates what is observed. When assessment is viewed narrowly as a descriptive device, it becomes a necessary piece of evaluation; somewhat of a lesser-included component. Yet, assessment can stand on its own, helping students to better understand their actions and where improvement is needed. For example, an evaluation can often be subdivided into two parts: observation (or assessment) and critique. Observation is a relatively disinterested objective description, reflecting the ability, skill level, or other quality of the person being evaluated, with a minimal subjective analysis. Critique includes judgment, which is a subjective analysis involving a comparison to others, an ideal standard, or a mythical average.

Evaluations often take the form of summative tests, which are likely graded and utilized as the sole basis for the course grade. Some evaluations, however, are diagnostic, simply determining a person's skill level or competency at a particular point in time. When taking a music or sports lesson for the first time, the instructor generally administers the student a diagnostic test to determine what type of teaching is appropriate. Some may view such a diagnostic test as an assessment, particularly if it lacks a judgmental or normative quality.

Evaluations in a formal education setting refer primarily to graded papers, exercises, or examinations.[33] A primary purpose of the evaluation is to measure and rank the students' skill levels, particularly their mental abilities.[34] Ranking is intended to reflect student achievement, but also to advance, on a larger scale, the underlying societal value of merit.[35] It is within the arena of meritocracy in which many tests, such as standardized tests, are framed.

Evaluations from primary school onward are utilized as a measuring device for intellectual ability. These devices measure present ability, such as performance tests, or potential

33. An evaluation may or may not be graded, depending on its function.

34. In law schools, the qualities assessed usually include legal analysis or "thinking like a lawyer."

35. *See* PETER SACKS, STANDARDIZED MINDS 5 (1999).

ability, such as many standardized tests.[36] The measurement of potential ability focuses on future competencies rather than on current ones.[37] In this way, a test serves as a predictor, not a "snapshot" of the present mental ability of the test-taker.[38]

1. *Examinations*

An examination is a type of evaluation defined by its usage: to measure particular skills or abilities.[39] An examination tends to be of limited time duration and it contains numerous instructions to guide the test-taker's responses. Formal and especially standardized tests are expected to provide an objective, reliable measure of the relevant skills being tested.[40] Grades yield either a percentile score relative to other test-takers or determine whether the test-takers have exceeded a minimum level of competency. The basic competency tests are often called criterion-referenced exams.[41]

Significantly, examinations testing mental ability are only representative of the skill or ability tested, whether it is general intelligence, understanding of the course material as a whole, or ability as an analytical thinker.[42] In this representative capacity, the examination serves as a predictor of how well the test-taker will do in another context. Hence, an examination is neither a simulation device nor an assessment tool that directly indicates either potential or present ability in an activity or

36. *See id.* at 27.

37. *See id.*

38. The predictive power of standardized tests, including the LSAT, is often less than believed by the public. *See id.* at 7.

39. According to several commentators, "grading is also used to recognize exceptional performance, to penalize unacceptable performance, and to provide employers with a way of distinguishing among students." Brustin & Chavkin, *supra* note 30, at 306.

40. Even if the test measures the appropriate skills, it must measure those skills consistently, or else it will yield distorted results. Further, the test must measure the same thing for all test-takers and measure material skills, not irrelevant factors.

41. *See* 1 Michael Josephson, Learning & Evaluation in Law School 4 (January 1984) (unpublished manuscript submitted to the Assoc. of American Law Schools Annual Meeting).

42. It is assumed that a law school examination captures the qualities of good analytical thinking. The question of what exactly good thinking is, and how it embodies analysis, creativity, or pragmatic problem solving, is not readily answered. Nor, for that matter, can it be said to even have been addressed by teachers professing to measure such "good thinking."

thing.[43] It also does not reflect past performance in a task or skill.

The use of examinations as a representative tool aptly characterizes law school evaluations, since the examination is not only considered an indicator of the quantitative and qualitative learning in a course, but how a student will perform as an attorney as well.[44] Further, the case method, which serves as the basis for testing, is a representative tool of both lawyering and critical thinking.[45] Law school tests, therefore, are intermediaries for lawyering competencies and thinking skills. As an intermediary, it is grounded on several salient assumptions. One assumption on which a law examination is based is that writing is inherently connected to thinking; such fuzzy writing is indicative of fuzzy thinking.

2. *Prerequisites to Examination Legitimacy*

The legitimacy of a test question[46] is predicated on three major requirements: validity, reliability, and efficiency.[47] Without these essential elements, the test will not be accepted as a proper measuring device and will lose its legitimacy.

a. *Validity*

A test is valid if it measures what it purports to measure.[48] This definition of validity is termed "content validity," although it has been called many other names as well.[49] Validity depends

43. *See, e.g.*, ROBERT STERNBERG & LOUISE SPEAR-SWERLING, TEACHING FOR THINKING (1996) (in which the authors consider what comprises "good thinking" and how teachers can teach thinking).

44. *See* Steve Sheppard, *An Introductory History of Law in the Lecture Hall*, *in* 1 THE HISTORY OF LEGAL EDUCATION IN THE UNITED STATES 7, 34–35 (Steve Sheppard ed., 1999).

45. *Id.* at 24–26.

46. A test question is also called a "test item." *See* Josephson, *supra* note 41, at 3.

47. *Id.* at 5–6 (Many scholars replace "efficiency" with "fairness." The two terms have the same practical meaning: that the test is "workable.").

48. *See id.* at 8.

49. *Id.* at 7 (noting that "Content validity focuses on the relationship between the subject matter of the test and the instructional objectives of the teacher," and that "an inappropriately designed test is like trying to measure light intensity with a barometer") (citing RICHARD LINDEMAN & PETER MERENDA, EDUCATIONAL MEASUREMENT 74 (2d ed. 1979)).

on the particular situation and its use.[50] It is not an all-or-nothing concept, but should be considered a matter of degree—such that the validity of a test could be low, medium, or high.[51] Validity is especially important to standardized tests. These tests, including the Law School Admission Test (LSAT), represent the ability to predict later success in school.[52] Statistical analysis is widely used to determine a test's validity.[53]

In the law school context, an exam is valid if it measures what the professor wanted the students to learn in the course, whatever that might be. For example, a written law school examination might test:

> (1) the student's knowledge of the subject; (2) the accuracy of the student's recall of the knowledge and his understanding of it; (3) how effectively and accurately this knowledge can be communicated; (4) how skillfully and efficiently this knowledge can be applied to particular circumstances; and (5) how rapidly these functions can be accomplished in an exam situation.[54]

If the test measures something other than what a professor intended students to learn in the course—such as particular do-

---

50. *Measurement, Reliability, and Validity*, at http://courses.smsu.edu/673f/lecture%20notes/chapter%205e.doc (last modified Feb. 19, 2002) (on file with Pace Law Review). Even if a test is valid and reliable, it must be accepted by the community in which it is given, otherwise its legitimacy will be undermined. This dimension of cultural acceptance evidences the social role of tests and their objective measurement. *Cf.* Daniel Gordon, *Does Law Teaching Have Meaning? Teaching Effectiveness, Gauging Alumni Competence, and the MacCrate Report*, 25 FORDHAM URB. L.J. 43, 62 (1997) (discussing alumni evaluations of their classroom experience as a training tool for practice). Competing demands include teaching, scholarship and committee work. If the true importance of evaluation is recognized, however, teachers might make room for more serious pursuits of evaluation creation and implementation. Further, teachers might be more aware of how to circumnavigate obstacles to anonymous and fair exams. *See generally* GERALD F. HESS & STEVEN FRIEDLAND, TECHNIQUES FOR TEACHING LAW 286 (discussing the importance of evaluation and providing students with feedback on their performance).

51. *Measurement, Reliability, and Validity*, *supra* note 50.

52. *See* Josephson, *supra* note 41, at 7.

53. The subject of psychometrics offers several statistical formulas for determining the validity and reliability of tests, such as the Spearman-Brown formula. *See, e.g.*, JIM C. NUNNALLY & IRA H. BERNSTEIN, PSYCHOMETRIC THEORY 212 (3d ed. 1994).

54. Derrick A. Bell, Jr., *Law School Exams and Minority-Group Students*, 7 BLACK L.J. 304, 305 (1981).

mains[55] of information or skills, then it is not a valid test.[56] The issue of validity can arise in several ways on law school exams.[57]

First, exam questions may prove invalid by focusing on material, issues, or information not covered in the class. These questions are out-of-bounds and exceed the permissible scope of a valid test.[58] Numerous problems arise with questions based on knowledge that a professor believes the class has acquired, but, in fact, the class has not. Questions utilizing current events, cultural references, or sports often may be invalid for some test-takers. For example, someone who is not a football fan may find little significance in a question involving football that speaks of a contract for a "one thousand yard rusher."[59] Similarly, a person who is new to the United States may find a question concerning Halloween tradition perplexing.

Second, a validity problem may reside in the "call" of the question. The "call" is the part of the question that asks the test-taker to do something, such as analyze the legal issues, make an argument on behalf of a party, decide who will win a case, or something else.[60] If the call of the question is unfairly ambiguous or vague, the test will not measure what the professor intended to measure. In many law school exams, the call of the question is broad and loosely defined, limited to "discuss the legal issues" or "explain the legal consequences."[61]

---

55. A domain of information is the particular subset, such as causation in negligence or mental state in murder, to be learned.

56. Often, law school teachers define courses by their substantive coverage. Unfortunately, however, there is no consensus among law school teachers or the legal profession as a whole about what the essay examination specifically measures.

57. See Professor Michael Josephson's fine discussion of validity on examinations. Josephson, *supra* note 41, at 6–14.

58. Questions that are based on material assigned, but not covered, lie close to the border of validity but still within it.

59. One student who stated that a football "rusher" was used as a character on one of her law school evaluations wondered why the person was in a hurry.

60. *See* Josephson, *supra* note 41, at 9.

61. *See id.* at 10. This lack of specificity could undermine validity if the test-takers are not appreciably aware of what the question asks. *See id.* at 9. Increased specificity minimizes problems with validity, but also tends to disclose issues. Since issue-spotting may be one of the skills tested, specificity and sub-parts in the call of the question may be disfavored.

Third, a validity problem may arise from the relationship among issues in an examination. If one issue is a gatekeeper for other issues it becomes overly important, skewing the likelihood that the test fosters an accurate reading of the test-takers' abilities. As described by Professor Michael Josephson, "[w]hen a single aspect of competence becomes an absolute condition precedent for success on an exam (rather than just one dimension of performance), the exam may be rendered invalid because the results become 'confounded.'"[62] This confounding creates difficulties not only for students who distort or miss the foundational gatekeeper issue, but also for graders who observe that some students perform well with the downstream issues but not the gatekeeper questions, and vice versa.[63]

Fourth, a related problem to linkages among issues is a failure to make a test representative of the content of a course. Spotlighting an obscure and unimportant issue on an exam, even if discussed briefly in a course, will skew the results by not testing what the students learned overall in a course.

b. *Reliability*

Reliability offers a somewhat different testing prerequisite than validity. It does not ask what a test purports to measure,[64] but whether such a test measures those skills or abilities consistently, from one test item to the next, or one test-taker to

62. *Id.* at 11.

63. Avoiding confounding or co-mingling of issues, is particularly difficult in issue-spotter exams that link together a series of events containing numerous legal consequences. In fact, sorting, distinguishing, and disconnecting issues are some of the skills that might be tested on the exam. To circumvent this problem, educational psychologists, according to Professor Josephson, suggest that

> tests include a variety of items to measure various types of competency at various levels of difficulty. Thus, a teacher who consciously desires to include a test item with a non-directive call ought to include additional test items with more specific calls so that the student can reveal other dimensions of legal competence.

Id. at 11. This approach analyzes from the perspective of the student, not the teacher, and reflects the substantial literature indicating that students learn and communicate that learning differently. The inference to be drawn from this supposition is that to most accurately measure skills learned in the course, a variety of test items will encourage the best performance from the greatest number of students.

64. That is, the test evaluates relevant skills or abilities.

another.[65] A reliable exam will measure like performances similarly,[66] whether it is comparing the similar performances of two different test-takers, of a single test-taker who has taken a test several times,[67] or even among individual test questions.[68] Thus, it is crucial to create a consistent grading mechanism. If a test is unreliable (and inconsistent), it will not be valid because it will not measure what it is supposed to measure.[69] In other words, an otherwise valid test lacking a reliable grading instrument is of little use.[70]

---

65. *See* Josephson, *supra* note 41, at 15. In this way, reliability is a necessary element of validity. If a test is not reliable, it will not measure what it purports to measure. *See id.* at 15.

66. *See id.* at 17–18.

67. *See, e.g.*, Ben D. Wood, *The Measurement of Law School Work*, 24 COLUM. L. REV. 224, 245 (1924).

68. Psychometricians use various mathematical formulas to test for reliability, such as the Spearman-Brown formula. *See id.*

69. *See* Josephson, *supra* note 41, at 15.

70. Item analysis:

1. Standard Deviations

Standard deviations measure the variability or distribution of scores. In essence, they reflect how spread out scores are for particular questions or examinations. A compressed standard deviation means that the scores from the test-takers are very similar and cluster around the average or mean. Likewise, a spread out standard deviation means there is a wide range of scores that don't cluster around the mean.

If test items with different standard deviations are merely added together, the item with the greatest standard deviation will have a greater impact on the result. Thus, if an examination has two equally weighted questions, the one with the greater standard deviation will be a greater determinant of the final grade.

2. Discrimination Index

The discrimination index is intended to reveal whether a test item was answered correctly more by good test-takers than bad ones. The discrimination index helps to determine whether the question was a good one, and for objective questions in particular, not excessively susceptible to guessing. The index ranges from plus one to minus one. A positive score means the question is a good one because more top test-takers answered correctly than poor test-takers. If a score is near zero, it means the test does not discriminate between good and bad test-takers. If a score is negative, it indicates there is a defect in the question, since poorer test-takers did better on the question than the good test-takers. For example, a discrimination index of 0.382 indicates that a question was answered better by the better test-takers, and was probably a proper question.

3. Point Biserial Index

The point biserial is another type of discrimination index. It reflects the difference between the number of top test-takers who answered a question correctly and the number of poor test-takers who answered the same question correctly.

1. *Inconsistency Problems*

Grading mechanisms may vary, ranging from a points system, which assigns points for spotting and discussing legal issues, to a system that emphasizes certain parts of an answer such as legal analysis, to a holistic review of the quality of an answer.[71] If the grading process is not replicated from paper to paper—or even within parts of a single exam—then the variability undermines the accuracy and legitimacy of the results.[72] Without accuracy, the test will not assess the mental abilities it seeks to measure.[73]

2. *Weighting Problems*

Another aspect of reliability involves how scores are totaled on exams. One problem regarding scoring occurs when standard deviations are different among test items, and the test

---

This assessment is helpful in determining whether a question is simply difficult as compared to being unfair. A difficult question ought to be answered correctly by a greater percentage of the better test-takers.

4. Mean scoring

The mean score is the average total test score based on the people who answered a test item—or the overall test—correctly. This average is useful in evaluating a class's performance.

For definitions of each of the above terms, see FOUNDATION COALITION, GLOSSARY OF TERMS *at* www.foundationcoalition.org/home/keycomponents/glossary.html (last visited March 31, 2003). For application of the above devices, see Dennis W. Field & Sheila E. Rowe, *Development and Analysis of a National Certification Exam for Industrial Technology,* J. INDUS. TCHR. EDUC., Summer 2001, *at* http://scholar.lib.vt.edu/ejournals/JITE/v38n4/field.html.

71. *See* Kissam, *supra* note 3, at 445–47. The holistic model, while subjective, has its fans. According to one commentator, this approach has the "capacity to take account of the skills of interpretation, conventional and creative imagination, practical reason, and practical judgment, all of which are associated with Aristotle's philosophy of ethical or normative decisionmaking." *Id.* at 446.

72. *See* Josephson, *supra* note 41, at 17–18. The openness to inconsistency provides an argument in favor of multiple-choice questions, since the grading process for multiple-choice is machine scored and at least consistent—treating all test-takers the same.

73. *See id.* at 15. There is no training, oversight or perhaps, even awareness, of the importance of consistency in law school evaluations. In fact, in contrast to the sociological benefit of publishing opinions by judges containing the rationales for their conclusions, the grading process generally occurs in an environment of total secrecy. Any disclosure is at the prerogative of the professor. *See* Kissam, *supra* note 3, at 445.

combines the scores of those test items.[74] This problem is a weighting error, in which certain parts of the examination are effectively weighted more heavily than other parts. Even if there are two equal parts to the examination, adding up the points of each part may yield a distorted score. Suppose one of the two parts of the exam yields almost equal scores for all and the other part yields a widely disparate set of scores. The part of the exam that will serve to rank and order the students will be the part with the wildly disparate scores. This part, then, will truly count the most. The part with the equal scores will count less. Thus, to make equal components of an exam count equally, adding up the scores when there are different point spreads or standard deviations just will not do. Instead, the teacher can curve each component of the exam, giving the same percentage of A's, B's, C's, etcetera, so that the variability in grading each component will not affect the weight of the test.

c. *Evaluation Efficiency*

A third evaluation prerequisite is not tied to the psychometrics of evaluation, but rather to the practical considerations of economy. The inquiry becomes: Will the administration and grading of an evaluation maximize outcomes while minimizing costs such as time and resources? Along these same lines: Will the use of in-class evaluation adversely affect other teaching goals such as the coverage of substantive material?[75] These potential costs often play a major role in determining the nature and type of examinations. The likely costs are directly relevant to decisions about law school examinations that are traditionally graded solely by the teacher—without assistance from teaching aides as in colleges and universities. Institutional

74. In fact, most professors feel compelled to weigh all questions equally. Ironically, such treatment still could lead to differing outcomes. Some professors total points earned on an exam, favoring students who reach all of the elements on a test. Other professors divide an exam into segments, allowing for excellence on certain parts of an exam to have a greater impact. For example, a student who gets A's on the first two parts of an exam and an F on the last part will still get a passing grade, whereas if the points were totaled, the student would receive a lower grade, perhaps even an F. In addition, a professor who uses the "gut" approach could value only the best part of the student's answers—effectively throwing out either a high or low aberration—in reaching a final grade.

75. This potential cost is a major impediment to the use of in-class evaluations in law school by many professors.

costs are also relevant. Schools usually do not give law professors credit for any time spent on grading or reviewing examinations. In fact, such review time often detracts from other professional responsibilities such as teaching, writing, or committee work.[76]

B. *Types of Law School Evaluation Instruments*

Law school evaluation occurs almost exclusively through written[77] final examinations,[78] using the classic issue-spotting form.[79] In seminars, final papers are substituted for timed examinations. In clinical courses, there may be some type of performance evaluation as well.[80]

1. *Essays*

Different types of essay test questions are used in legal education,[81] with the traditional issue-spotter predominating.[82]

---

76. *See* GREGORY S. MUNRO, OUTCOMES ASSESSMENT FOR LAW SCHOOLS 163 (2000).

77. There are many kinds of evaluation other than written tests. These include oral tests, classroom contributions, and written reviews of performance. Oral examinations, for example, are used in some graduate programs such as in medical and dental school. *See* HEYWOOD, *supra* note 6, at 40–42.

78. "The majority of American law schools use a single, end-of-term written essay exam to assign grades." Douglas A. Henderson, *Uncivil Procedure: Ranking Law Students Among Their Peers*, 27 U. MICH. J.L. REF. 399, 403 (1994).

79. Professor Kissam observed: "The first of these examination functions is issue spotting or, more precisely, perceiving analogies between the stated facts of an examination problem and professionally recognized legal issues, standards, and precedents." Kissam, *supra* note 3, at 440.

80. Performance testing is becoming more popular, as exemplified by its inclusion on many state bar examinations. *See* SOC'Y OF AM. LAW TEACHERS, STATEMENT ON THE BAR EXAM, *at* http://www.saltlaw.org/positionbarexam.htm (July 2002).

81. Similarly, there are different types of answers to essay questions. For example, one commentator has offered five different types of answers to essay questions, including (1) the "key sentence" answer, in which the test-taker correctly identifies the important facts and the relevant law; (2) the fact omission answer, which omits a key fact necessary for resolution of the item; (3) the law omission essay, in which the student omits the applicable law in analyzing the question; (4) the conclusion, where the test-taker correctly identifies the answer but without analysis of the facts or law; and (5) the irrelevant response, where the writer discusses an irrelevant theme instead of the pertinent question. *See* Lawrence Vold, *Types of Essay Law Examination Answers—Good and Bad*, 3 HASTINGS L.J. 85, 85-90 (1951).

82. *See* Norman Redlich & Steve Friedland, *Challenging Tradition: Using Objective Questions In Law School Examinations*, 41 DEPAUL L. REV. 143, 143 (1991).

This dominant essay type asks students to first identify the issues lying within a fact pattern and then to analyze them.[83] Other types of essay questions include (1) the critique question, which asks students to criticize or justify a rule of law, case, doctrine, or rationale questions, and (2) the role-play essay, which asks students to take the side of a party or play a judge in resolving an issue or case.

The traditional issue-spotter essay question is intended to determine how well the test-taker "thinks like a lawyer."[84] This is one of the major skills covered in the first year of law school.

2. *Short Answers*

Instructors often use short answer questions as directed essays, asking students to respond to a specific issue with time or space constraints. The short answer question permits broader coverage of issues while maintaining the importance of the student's writing—i.e., communication skills.

3. *Selected Response*

Selected response questions ask students to choose a response from a finite number of choices. The most common type of selected response question is multiple-choice, usually including four responses after a call of the question. There are different kinds of multiple-choice questions, from "choose the best/worst answer" to "choose the applicable rule of law/proper legal analysis."

Selected response questions usually take from one to two minutes to complete. On the Multistate Bar Examination, students are given 1.8 minutes per selected response question.[85] Because they are quickly answered and easily scored, this type of question can test a wide number of areas quite readily.[86]

---

83. The issue spotter essay has been described as requiring students to detect analogies "between the stated facts of an examination problem and professionally recognized legal issues, standards, and precedents." Kissam, *supra* note 3, at 440.

84. Henderson, *supra* note 78, at 399.

85. The Multistate Bar Examination (MBE) is a two hundred question multiple-choice examination lasting six hours. ABOUT THE MULTISTATE BAR EXAM, *at* http://www.thebarexam.com/aboutmbe.htm (last visited Dec. 27, 2002).

86. As one commentator noted: "The greatest strength of an objective examination is its ability to test a great number of narrow areas." Howard J. Gensler, *Valid Objective Test Construction*, 60 ST. JOHN'S L. REV. 288, 289 (1986).

4. *True/False*

True/false questions present an assertion and ask the test-taker to choose whether that assertion is either true or false. Generally, no explanation of an answer is permitted. This test item is the most susceptible to guessing, since the test-taker has a one-in-two chance of reaching a correct answer.

5. *Guided Essay*

The guided essay combines the multiple-choice and essay formats. The tests feature a series of multiple-choice questions. However, in addition to choosing the best answer, students are asked to explain their choice.

C. *Upon Closer Scrutiny: The Power Of Evaluation In Law School*

While occupying only a small space in the entire legal education enterprise, evaluation exerts significant influence over the whole law school educational process and community. While evaluation is believed to be a gatekeeper of admitting the best students to the schools and the profession, it really is a gatekeeper of another sort—the overall nature and organization of the law school experience.

Upon closer scrutiny, its power is clear. Evaluation dictates who will be a successful law student, eligible for all of the vestiges of that success, and who will not. Yet, evaluation, because of its interconnectedness, also casts a large shadow over the entire learning process, creating incentives for students to behave in a certain way and even serving to construct the law school experience. A critical assessment of the evaluation process serves to illuminate the pervasive influence of the evaluation process, as well as how evaluations can be better utilized.[87]

1. *Evaluation as Authority and Control*

The final examination at the conclusion of a course almost completely defines the entirety of law school performance, espe-

87. As one commentator generally noted about evaluating the social impact of examinations: "After all, it will scarcely be possible to regulate testing effectively unless its workings and consequences are well understood." HANSON, *supra* note 1, at 7.

cially in the first year of school.[88] While there are exceptions, the final examinations invariably provide the bulk of a student's grade. This indubitable fact has many consequences, not the least of which is its profound effect on how students prepare for and learn in a course.

On a structural level, the final examinations shape the law school ethos—the unstated set of axioms and postulates that govern the law school experience. Rather than being a place to learn a profession, the dominance of examinations removes incentives to treat the experience as a professional gateway, and instead replaces that with the incentive to improve on test performance. The continuation of an examination orientation signifies to students it is "schooling as usual," rather than a pragmatic and active learning environment. Such an approach encourages conformism and individualistic achievement.

The importance of final examinations encourages certain types of learning processes and shifts the emphasis on what students learn and why. Students write outlines of the material covered in their classes to better understand a course, and more specifically, to better write out essay exam answers. Students purchase a wide array of study aids, with the aim of not just understanding a course better, but understanding it better for an exam. Students, perhaps naturally, become competitive with each other over the perceived race for good grades,[89] or consciously resist the tendency to do so. Students focus on what might be included by the professor on a test, paring away interesting but unexamined sidelights. The question, "Will this material be on the exam?" often lurks just below the surface and sometimes even percolates to the top when students are confused about course coverage or their responsibility for the purposes of examination.

Learning, by force of the exams and perhaps by choice, is entirely instrumental. Learning is channeled into the preparation for examinations. Students, especially those who have ex-

88. Evaluations do not define everything. There are exceptions—moot court, A.T.L.A. membership, clerking, and pro bono opportunities may lie outside of the influence of evaluations. A law school's evaluation culture, however, can be defined in large part by how the school approaches grades.

89. Some students believe, even if it is not true, that classes are graded on a curve and that there are only a limited number of good grades available in any one course.

perience in taking law school examinations, often switch preparatory habits, forsaking those that emphasize class readiness to those that promote exam readiness.

While class is an opportune place for dialogue and analysis, it becomes apparent to students that creativity beyond the exam is irrelevant at best and even deleterious to the assessment of performance. Voices of rebellion or anti-authority may be expressed in class, on e-mail, or around the school, but exams have a way of lowering the boom—such voices likely would not receive credit and may even be considered insurgent.

In class, the shadow cast by examinations shifts attention to the teacher's agenda and away from students making comments. Indeed, there are students taking copious notes of everything a teacher says at the expense of student comments, who are relegated to play the role of foils.

Ironically, the influence of evaluations extends to class attendance and participation. While some professors offer attendance inducements,[90] class attendance may be sustained by the belief that it helps on a particular professor's examination. Conversely, classes where it is perceived that the class experience offers little assistance on an examination may prove to depress attendance. Furthermore, examinations can influence class attendance at the end of the semester. Students often skip end of the semester classes to study for examinations. This effect diminishes the importance of class attendance.

Class participation is also indirectly affected by evaluation. When students realize after their first semester of school—and the receipt of first semester grades—that quality of participation may not correlate with the grade received in a course, the student may be less inclined to participate. On the other hand, students who receive confirmation of their "legal intelligence" from good grades may be more inclined to participate.

The control exercised by exams is perhaps at its zenith when it comes to eligibility requirements for student organizations and law-related jobs. Student-run journals, such as law review, moot court and other organizations, often require cer-

90. In some courses, the professor will offer extra credit or exercise a grade reduction based on class participation. Grade reductions may occur for failures to attend class as well.

tain grades to even try out for the organization. Some firms, when interviewing for summer or permanent associates, will only choose from students with the highest grades. Desirable judicial clerkships are given to the students who have performed the best on examinations.

a. *A Student Survey*

A spring, 2000 student survey tested many of these assertions concerning the authority and power of examinations.[91] This survey, while small and admittedly unscientific,[92] provides some illumination on student motivation and behavior. When asked whether they learned the materials in the first semester of law school because they were truly interested in the subject matter, eighty-four percent of the respondents agreed or strongly agreed with such an assertion, while only sixteen percent disagreed or strongly disagreed with it.[93] When asked the same question of their second year of law school, only seventy-two percent agreed or strongly agreed that they learned the materials because they were truly interested in learning the subject, while twenty-four percent disagreed or strongly disagreed.[94] When asked whether their approach to exams changed from the first semester to the second semester of law school, seventy-six percent of the students stated that it had indeed changed.[95] Changes also occurred in the way students approached their classes, including attendance.[96] While these changes cannot all be attributed to examinations, there is some evidence that the evaluation process did play a role in the stu-

91. This survey occurred informally and included students at the Nova Southeastern University Law Center [hereinafter NSU Law Survey] (on file with the author).

92. Twenty-five students from Nova Southeastern University responded to the questionnaire. Fourteen of the students were in their first year of law school and the remaining students were in either their second or third year of law school. *Id.*

93. *Id.*

94. *Id.*

95. *Id.*

96. One student stated: "I quit attending regularly because I was disenchanted with my grades given the effort I expended." Yet another student stated: "My exams affected my attendance in class in that the exams reinforced and motivated me to attend." NSU Law Survey, *supra* note 91.

dents' attitudes and motivations toward their law school careers.

2. *Social Construction and Evaluation*

The law school experience is subject to social construction. It can be viewed[97] as created by groups, such as the experience of first-year sections, through taking all of the same courses, and by particular teachers. The social construction is not linear or chronological, but based on impact and perception. It can be seen in the spirit of the school as part of a larger university or from the tenor set by the dean.

In light of its disproportionate power, the social construction of law school can be seen as revolving around the evaluation process. Although treated institutionally as an afterthought, the perception among students, alumni, and employers alike focuses attention and importance on the evaluations, not what precedes them.

While these perspectives play a large role in a person's law school experience, one commonality is the evaluation process. In fact, evaluation often constructs much more of a student's experience than just the time spent preparing for and taking examinations. The examinations not only serve as notice of performance, but they also rank students, determine which students get opportunities to join limited membership organizations, and indicate which students are considered to be a success overall. The impact extends from prestige and potential jobs, to the psychological, bearing on self-esteem and self-image.

The evaluation creates the most important school hierarchy—class rank—and defines who will be on law review and the other law journals, who is eligible for jobs with large law firms

97. Perspective counts when assessing legal education. Much like the film *Rashamon*, which demonstrated the power of perspective, students have a very different approach to class than professors; professors have a different approach to the school than administrators; and administrators have different perspectives on the law school experience than librarians. In the book *One L*, for example, first-year law student Scott Turow described his first demanding and nerve-wracking year in law school, with perhaps the most harrowing time period being that of exams. SCOTT TUROW, ONE L 173–99 (1977). Others experiencing the first year of law school express wonder at how their experience and Turow's could be so different.

and judicial clerkships, and much more. While some student organizations, such as moot court and ATLA (American Trial Lawyers' Association) may not require grades for admission, some do have "mandatory minimums" for participation. Even research positions with faculty members often depend on grades.

The constitutive power of the first-year grades can have a far-reaching measure of social stimulus and control.[98] In that first year, a student's grades may have led to a position on law review, a judicial clerkship, and then a position with a large firm or competitive governmental agency, placing that person in a position for further advancement. A student who did poorly in the first year—and perhaps very well in the third—may have few of these options, and must obtain entrance to these positions through alternative paths.

Similarly, courses that are not graded often reflect a lower status in the law school culture. As two commentators have observed, "[t]he institutional decision to grade clinical courses on a pass/fail basis reflects the view of some non-clinical faculty that clinical courses are less intellectually demanding."[99] On the other hand, it also could reflect an inability to create adequate standards for evaluating student performance in such courses.

There are exceptions, of course, to the constitutive power of evaluation. Inter-school moot court and ATLA competitions often are held in high regard, and some students who participate in student affairs, politics, or pro bono activities are noticed and rewarded for their efforts. However, the students with the highest grades are singled out at graduation and have the options accompanying such status.

A more protean effect of law school examinations is its impact on local law school culture. How students and teachers create an environment around evaluations is telling. Institutionally, some schools police the examination process exten-

98. As one commentator noted about testing in general: "If a decision-maker can point to the results of an objective and valid test as the information on which a control decision was based, those being controlled are more likely to accept and internalize the decision and its consequences." ANTHONY J. NITKO, EDUCATIONAL TESTS AND MEASUREMENT: AN INTRODUCTION 39 (1983).

99. Brustin & Chavkin, *supra* note 30, at 307.

sively, do not release grades publicly, and laud students who achieve higher grades (while ignoring those who do not). Other schools de-emphasize grades,[100] downplay its hierarchical value, and work extensively with students who may be educationally disadvantaged in academic assistance programs.

Students also may choose to adapt to grades and evaluations differently. In some schools, there is significant competition, with little sharing of materials and stories about mutilated or missing books in the library. In other schools, students have an outline bank that collects and disseminates outlines for the student body's use, and students have an encouraging and cooperative attitude. The local culture can change from year to year, entering class to entering class, and even section to section.

3. *The Politics of Evaluation*

A significant element of the control exerted by law school evaluations is arguably of a political nature, tied to the role of evaluation as the gatekeeper to the profession and to the schools themselves. The use of evaluation for political purposes extends back decades. For example, one of the champions of intelligence testing, Charles Spearman, wrote in 1927 that "an accurate measurement of every one's [sic] intelligence would seem to herald the feasibility of selecting the better endowed persons for admission into citizenship—and even for the right of having offspring."[101]

While discussions about the exclusionary role of testing have become more muted, many suggest that such claims still operate in a sub-rosa fashion. Arguments about the political role of law school evaluation also exist. Some people claim that the evaluation process is not objective,[102] but instead operates in a biased manner against minorities and groups without

100. YALE LAW SCHOOL, JD PROGRAM: ACADEMIC REQUIREMENTS & OPTIONS, *at* http://www.law.yale.edu/outside/html/Academics/acad-jd.htm (last visited Oct. 31, 2002).

101. CHARLES SPEARMAN, THE ABILITIES OF MAN 8 (1927).

102. *See, e.g.*, Crane, *supra* note 3, at 791 ("The 'objectification' of grading essay examinations is [sic] also allows law professors to compensate for the inconsistency in the way they test the subjects and the way they teach the subjects.").

power.[103] It is asserted that some tests, such as the LSAT, are culturally biased and written in such a way as to prejudice minority test-takers.[104]

Given the significance of the law school examination, objectivity appears to be an essential component of a fair evaluation process. To promote this fairness, many schools utilize anonymous student examination numbers. Furthermore, many individual teachers use grading checklists, grading each question on a point-by-point basis and distancing themselves from a subjective hunch approach.[105]

III. AGING MONUMENT STATUS: DEFECTS IN THE EVALUATION PROCESS

The supplementary status of evaluation[106] has resulted in numerous problems, from overvaluation as a measuring device to undervaluation as a teaching strategy. Corollary downstream problems have occurred as well, such as a pervasive influence of examinations throughout the entire law school experience.

While the problems emanating from the evaluation process cannot all be attributed to the institutional stance toward evaluation, a faculty and school's policies undoubtedly contribute to the current status. The initial and perhaps foremost problem is the institutional marginalization of evaluation relative to other components of legal education.[107]

---

103. *See generally Legal Issues In Testing*, *at* http://ericae.net/db/edo/ED289884.htm (last visited Sept. 9, 2002) ("The cases are usually based on the argument that the tests are biased against the lower scoring group or that they reflect the effects of past segregation in the schools.").

104. *See id*; Bell, *supra* note 54, at 306–07.

105. One commentator called this "objectifying" the essay examination. The commentator claims that this objectification reduces subjectivity and increases the fairness in grading. *See* Crane, *supra* note 3, at 788.

106. The supplementary status of evaluation in the constellation of legal education results from a variety of causes, including a lack of institutional oversight, an undervaluation of its value as a pedagogical tool, and an overvaluation as a measuring device. Further, history, tradition, a lack of faculty training, and job-related disincentives contribute to the lack of importance of evaluation.

107. This lack of oversight, when combined with a lack of training and low priority for the law teacher, leads to numerous problems. Many lawyers and law professors have their own favorite law school "examination nightmare" stories. For example:

A. *Institutional Marginalization*

Historically, American law schools have seldom accorded the evaluation process significant status.[108] Law schools became a sustained part of the environment in the United States during the Industrial Revolution of the late 1800s.[109] Christopher Columbus Langdell, a Dean at Harvard University, essentially invented the casebook in the early 1870s,[110] giving rise to an easy and uniform method of university-affiliated legal education. Prior to that time, most attorneys still entered the profession through an apprenticeship, rather than formal schooling.[111] The apprenticeship, ironically, was filled with ongoing evalua-

---

A student (now a law professor) was confronted by a question concerning a football scenario about a successful rusher. The student, lacking familiarity with the game of football, wondered what the player was doing "rushing" around all over the place. Professor Paula Lustbader, Remarks at the A.A.L.S. New Law Teachers Conference (June 27–28, 2002).

In a core first-year course take-home, the teacher "borrowed" a question from another professor's prior examination. That question, in part, had been used to illustrate a model question and answer in a commercial outline, to which some of the students had access.

The teacher devoted a large portion of the test to an obscure footnote in the textbook barely touched on in class.

A professor administered a surprise in-class quiz, covering material outside of the assignments, and catching even completely prepared students off-guard.

The professor graded the final exam several months after it had been given, so that students did not receive the grades until well into the succeeding semester.

Typographical errors changed the nature of the question and the substance of the students' answers.

These and other stories about examinations abound. The stories are not surprising, given that a central, clearly articulated and serious role for evaluation in legal education is a far cry from the current stasis.

108. *See* Steve Sheppard, *An Informal History of How Law Schools Evaluate Students, with a Predictable Emphasis on Law School Final Exams*, *in* 2 THE HISTORY OF LEGAL EDUCATION IN THE UNITED STATES 815, 816 (Steve Sheppard ed., 1999). Of course, this marginalization can be contrasted with the historical preoccupation with intelligence testing. Some significant examples include literacy tests for voting, IQ tests, and the eugenics movements. *See* SACKS, *supra* note 35, at 48–49.

109. *See* Daniel R. Hansen, *Do We Need the Bar Examination? A Critical Evaluation of the Justifications for the Bar Examination and Proposed Alternatives*, 45 CASE W. RES. L. REV. 1191, 1197 (1995). "Prior to the Civil War . . . [i]f a university or local bar initiated a law school or formal legal educational program, it always failed within a few years." *Id.*

110. *Id.* at 1198.

111. *Id*; *see also* ROBERT STEVENS, LAW SCHOOL: LEGAL EDUCATION IN AMERICA FROM THE 1850S TO THE 1980S at 24 (1983).

tion and critique.[112] In this manner, the evaluation process was central to learning in an apprenticeship environment. The law schools failed to similarly recognize the importance of critique.

In law schools, a tradition of final exams as the sole means of evaluation developed in the late 1800s to help employers distinguish between students.[113] In the early 1900s, Professor Ben Wood wrote about this tradition and its defects.[114] Examinations were used to measure and motivate students, providing a negative incentive to perform or be dismissed.[115]

Schools rewarded faculty members for their scholarship and teaching, but incentives rarely, if ever, included any relating to the evaluation process.[116] Indeed, law school evaluations occurred in relative darkness, producing illumination only in the form of mysterious grades allegedly correlating with law student performance.[117]

Few resources were devoted to the evaluation process, other than perhaps proctors. Traditionally, there has been no formal training in the construction of evaluations or in grading those evaluations,[118] and there have been few, if any, resources devoted to the evaluation process by law schools and universities.[119]

While the number of law students in American schools has grown exponentially, and new buildings and facilities have proliferated, evaluation rarely has been important to the development or advancement of either the faculty or the institution. The U.S. News and World Report ranking of law schools, and the employer or public opinion of them, ignores the school's evaluation process.[120]

---

112. *See* Sheppard, *supra* note 108, at 817.

113. *See* Kissam, *supra* note 3, at 464–65. One commentator noted that the 1880s brought significant change as a result of New York City law firms desiring to see how law students performed on final "Blue Book" examinations. The greater the pressure to distinguish between law students, the greater the pressure to "objectify" the evaluation system. *See id.* at 464–65.

114. Wood, *supra* note 67.

115. *See* Henderson, *supra* note 78, at 402.

116. *See* MUNRO, *supra* note 76, at 163.

117. *See* Kissam, *supra* note 3, at 448–49.

118. *See* MUNRO, *supra* note 76, at 163.

119. *See id.* at 157.

120. *Best Graduate Schools 2003: Law, Methodology*, U.S. NEWS & WORLD REPORT, http://www.usnews.com/usnews/edu/grad/rankings/about/03law_meth.htm.

The institutional marginalization is reflected in the viewpoints of many individual teachers as well. The instructors obtained their positions by succeeding on timed essay examinations. Thus, it is no surprise that these are the types of exams that instructors draw up. Furthermore, school officials do not evaluate instructors based on the tests they give (teaching evaluations occur prior to exams).[121] Because the teaching obligation is more immediate than demands relating to the examination, professors often draw up the exams at the end of a course, possibly without extensive deliberation and without efforts to ensure continuity or proportionality. The fact that school officials consider the creation and grading processes insignificant to a professor's career advancement underscores the lack of immediate importance of examinations. (In essence, the teachers facilitate the cabining of examinations away from the substance pedagogy of the course). Instructors do not complete grades until several weeks after exams and students sometimes do not receive the results until well after the test is over.

Institutional marginalization has many consequences, some of which have already been noted. Among the consequences is the lack of standards that results from a lack of oversight. Without attention or rigor, examinations are accorded *carte blanche* status as a measuring device. When evaluations are used as a measuring device, particularly final examinations, it must be assumed that the examination will test the appropriate legal abilities and be fair and reliable.[122] However, this assumption is not a viable one.

(last visited Dec. 27, 2002) [hereinafter U.S. NEWS & WORLD REPORT]. However, administrators may indeed be aware of law school evaluation since law school grades may be a potential indicator of how students will perform on the bar exam. These administrators may recognize that the bar passage rates factor into the ranking of law schools.

121. Some two-semester courses in the first year have evaluations at the end of the first semester. If this is the case, however, these responses may be muted because students know the professor will grade them at the end of the second semester. Further, the first semester evaluation probably occurs prior to the examination.

122. Students do not often share this institutional assumption. *See, e.g.*, Steven H. Nickles, *Examining and Grading in American Law Schools*, 30 ARK. L. REV. 411, 443–44 (1977) (the author reported the results of a nationwide questionnaire on grading and assessment, and concluded that law students "are not satisfied with the examination process"). *Id.* at 439.

B. *Overvaluing Examinations As a Measuring Device*

Unlike law students, who often come to believe that "[g]rades [are] almost totally arbitrary—unrelated to how much you worked, how much you liked the subject, how much you thought you understood going into the exam, and what you thought about the class and the teacher[,]"[123] many in the law school community believe final examinations fairly and accurately assess a student's potential competence. Yet, there is no evidence that exams do so.[124] Even with standardized tests, to which all students must submit, psychometrics indicates these exams have much less predictive value than people would like to think.[125] In one sense, the tests really measure a particular thinking style,[126] one that exists within a world of multiple-choice questions and best answers.[127]

These assumptions do not survive careful scrutiny. Rather, the current examination process suffers from defects ranging from the philosophical to the pragmatic. As one commentator lamented, the single essay final examination is invalid, unreliable, and even "anti-educational."[128] Another commentator rejected the ranking system because it "devalues human beings, distorts the legal system into a cultural compactor, and diverts scarce resources from truly legitimate educational goals."[129]

---

123. Henderson, *supra* note 78, at 399 (second alteration in original).

124. *See, e.g.*, J.A. Coutts, *Examinations for Law Degrees*, 9 SOC'Y PUB. TCHRS. L. 399, 402–03 (1967). Professor Coutts states that the "educationalists have so far told us comparatively little" about how to evaluate the purported goals of analysis, synthesis and evaluation, and that the tests instead assess the lower levels of knowledge, comprehension, and application. *Id.* at 403.

125. SACKS, *supra* note 35, at 1–3.

126. Professor Kissam argues:

> The widespread belief among professors that Blue Book exams test for integrational ability, constructive thought, and writing ability does suggest a deeper truth about the style and content of Blue Book answers. This belief may reflect an inchoate and ill-defined perception of an implicit paradigm of successful Blue Book answers, a paradigm that I refer to as "good paragraph thinking."

Kissam, *supra* note 3, at 443.

127. SACKS, *supra* note 35, at 201.

128. Christopher T. Matthews, *Sketches For a New Law School*, 40 HASTINGS L.J. 1095, 1104 (1989).

129. Henderson, *supra* note 78, at 400.

To some, testing "mental ability" is itself deceiving.[130] Since mental ability is something not discerned physiologically, it must be ascertained only on a representative basis. In this regard, it is not objective or even a performance assessment, like a footrace, but a subjective substitute.[131]

The harm from potential testing comes in the form of the message it sends. What these tests purportedly reveal is how far students have to go to be complete, not what they can do right now.[132] In essence, they examine what a student will not be able to demonstrate, as much as what in fact they know—and can perform—at exam time.[133]

An interesting corollary to the use of tests to determine admission is their role in determining and confirming the reputation of a school. The school that takes the best students euphemistically means the school that takes the students who score the highest on standardized tests. The rankings of law schools, especially the influential U.S. News and World Report rankings, weigh heavily the entering LSAT scores of students,[134] however, the tests serve a different barrier purpose of establishing elitism.[135] Significantly, the distinctions in test-taking ability often fall along socioeconomic lines.[136]

1. *No Training or Support in Creating and Grading Evaluations*

The lack of training in the creation of valid and reliable examinations[137] contributes to the overvaluation of examinations

130. *See* SACKS, *supra* note 35, at 201.

131. A performance assessment actually determines what students can do on relevant matters as of the time of the examination. Objective multiple-choice questions, or even timed essay examinations, are often not relevant to lawyering. *Id.* at 201–02.

132. *See id.* at 232.

133. *See id.* This negative approach eschews the "merit badge" approach of the scouts and adopts a "distance from the finishing line" orientation instead.

134. U.S. NEWS & WORLD REPORT, *supra* note 120.

135. *See* SACKS, *supra* note 35, at 2. This argument is not accepted by all, but is by all means provocative.

136. *See id.* The U.C.L.A. law school admissions process exposed these socioeconomic disparities. *See id.* at 299–300.

137. Significantly, there is no training for law teachers on the teaching component of the profession. While the one and one-half day conference for new law teachers offers some insights into teaching, law teachers are less qualified and trained than almost any other type of teacher. *See supra* note 12. Indeed, elemen-

as a measuring device. The only constraints placed upon teachers often are administrative in nature, such as when they must administer the exam, when the test must be ready for copying, or the time limit for turning in grades to the school. In addition to a lack of training, there are rarely any institutional safeguards in place to ensure validity or reliability in the administration of exams.

Instead of devoting extra attention and resources to evaluation, befitting its power, institutions exert pressures on professors to minimize the time spent on the evaluation process. Given this history, the creating, administering, and grading of evaluations are not a traditionally important endeavor of serious teachers and scholars. In recent years, it has also been squeezed by the premiums placed on alternative activities and interests, such as promotion, tenure, and vacations. Time spent on evaluation is time that could be spent on the more highly rewarded activities of teaching and scholarship.

2. *Defects in Validity*

"[W]e teach these young [football players] how to compete and think under pressure. Those are things that make a person a true success."[138] —Vinnie O'Connor

It is widely assumed that law school examinations, particularly the single final examination, test what they are supposed to test—law student performance[139]—by effectively critiquing

---

tary school teachers are much better trained and schooled in the fundamentals and practices of good teaching than law professors.

138. Sal Ruibal, *Love Beats Lack of Money*, USA TODAY, Nov. 2, 2000, at 1C (quoting Vinnie O'Connor). Vinnie O'Connor has been a football coach for forty-seven years at St. Francis Prep. *Id.*

139. Given the importance of the examination process, it may be relevant to pose the question of how a knowledgeable person should evaluate law school evaluations. If some assessments in law school are better than others regarding their educational utility or their fairness, then this question is at least appropriate to ask. Of course, even at first blush, the response to this question does not appear to have a bright-line answer. Instead, any reply seems to depend on a variety of factors, including who is doing the reviewing, the goals of the professor for that specific course, and the general educational philosophy of that professor, among others. It's like asking which building is better when the answer depends, at least in part, on the building's intended use. Further, designing administrable evaluation review criteria may be as difficult as creating the building or examination instrument in the first place. For example, are some teaching goals preferable, such as favoring coverage of subject matter over lawyering skills? If it is possible

whether students "compete and think under pressure."[140] Yet, this thesis is for the most part completely untested and unknown,[141] especially as the practice of law changes in an age of technology and varying demands. For example, it is assumed that students have had notice of the evaluation process, in terms of its content and its locus at the end of a course, and consequently come ready to "compete." Yet, this assumption may be quite wrong. First-year law students must adjust to the relocation of the evaluation process,[142] from regular and periodic testing, to the pressures associated with a single test determining the entire course grade.[143] Without quizzes and the accumulation of grades, students must accept the ambiguities and suddenness of a single, final grade. Students must also adjust to the singularity of essay examinations, often containing only the most general of directives such as "discuss the legal issues."

The contention that law school exams test whether students "think like a lawyer" begs the question of what a law student should be learning in law school. That law school exams measure the "right stuff" is all but assumed. Yet, is there but a single lawyering competency, or many measures of excellence? Do lawyering skills have clear representative competencies, which the timed essay exams assume, or are competencies re-

---

to develop a hierarchy of goals, or if it is determined that all goals are considered equally valid, could the testing of some objectives, such as performance-oriented goals, be accomplished in a similar manner to that of other objectives, such as information delivery goals? These questions may have no clear answer. Raising them, however, may yield some valuable observations.

140. *Id.*

141. One law professor argued that:

> There is a remarkable inconsistency between the method of teaching law school subjects by focusing on judicial opinions, which often consists of analyzing "pivotal law or test cases," and the method of testing law school subjects by using borderline fact laden hypothetical situations that are designed to test the students' ability to reason or to perform "legal analysis."

Crane, *supra* note 3, at 786 (citation omitted).

142. "For most students, the curriculum before law school featured evaluation through mid-terms, class presentations, take-home exams, group projects, individual feedback, and term papers. Virtually none of these, however, survive in the modern law school . . . ." Henderson, *supra* note 78, at 399.

143. The administration of a single post-course examination often shocks first-year law students who are used to having multiple evaluations—tests, quizzes, and papers—during a course.

ally conventions varying across time and place? As the highly publicized MacCrate Report noted, law schools are doing too little to inculcate students with the wide variety of skills a lawyer uses in practice.[144] If a law student is an excellent negotiator but weak with case comparisons, why should the profession not find that this individual meets the requirements for practice?

a. *The Problem of Under Testing*

Arguably, exams measure only a limited number of skills relevant to legal practice, omitting much of what makes practicing attorneys successful.[145] As one commentator noted, limitations may occur as much for pragmatic reasons as for philosophical ones: "The controversy, uncertainty, novelty, originality, and practical judgments that are involved in both constructing and grading evaluative arguments [including judgment] lead most modern law professors to avoid such questions in order to ensure 'objectivity' in grading."[146] Thus, law school essay tests are of questionable validity in predicting future competence as a lawyer or in measuring important lawyering skills.[147] The timed essay may not predict future competence because few, if any, attorneys are measured by how well they analyze a legal problem where the facts are already digested and presented to them, a response must be given in minutes, and no references, research, or follow-up questions are permitted. For the same reasons, the essay may not serve to measure important lawyering skills, such as research, fact organization and presentation, creation of a theory of the case, and even legal analysis.[148]

---

144. AMERICAN BAR ASSOCIATION, LEGAL EDUCATION AND PROFESSIONAL DEVELOPMENT – AN EDUCATIONAL CONTINUUM, REPORT OF THE TASK FORCE ON LAW SCHOOLS AND THE PROFESSION: NARROWING THE GAP 5 (1992) [hereinafter MACCRATE REPORT].

145. In a study in the early 1920s, Professor Wood of the Columbia University Law School found that essay examinations did not validly measure overall legal ability. Wood, *supra* note 67, at 225–26.

146. Kissam, *supra* note 3, at 442.

147. *See* Redlich & Friedland, *supra* note 82, at 146.

148. *See* Henderson, *supra* note 78, at 409:

> Virtually all law professors agree that the law school essay only tests a limited set of skills, namely those relating to "thinking like a lawyer." Precisely because "thinking like a lawyer" does *not* represent all of the skills needed to

b. *Under Tested Domains—Ethics and Skills*

Law schools have made great strides in teaching ethics (and its corollary, judgment) in recent years. However, evaluations have relegated ethics to a secondary, less respected subject. Ethical components of basic courses are often left untested, sending the message that these issues are of secondary importance. Further, the subject matter is generally relegated to the contents of a single upper-level course, with few credits and subject to an examination like any other class. Thus, it lacks the primacy of a multiple credit first-year class or the uniqueness of a simulation course.

After a lengthy investigation, the MacCrate Report recommended that more lawyering skills be taught as part of legal education.[149] Many schools responded to the study by adding more clinical-oriented courses and opportunities for students.[150] However, the report failed to confront a major impediment to such changes—the ability of law schools to properly test those lawyering skills. How to measure skills such as client counseling, negotiation, and oral advocacy has perplexed many in legal education, particularly those with an essay "mindset." Even if such skills could be readily tested, if the evaluations were untraditional the results probably would be suspect and treated with skepticism.

There are intrinsic problems associated with a single event to mark progress and potential, as compared to a multiplicity of evaluations. In essence, the essay final examination fails to provide helpful information about a person's skill trajectory regarding their degree of improvement over time.

Law school faculties do not ensure the validity of law school exams. That is, the faculties generally do not "define the domain or tasks to be measured . . . analyze the mental process required by the tasks . . . compare the scores of known groups

---

practice law, the standard legal test provides an incomplete measure of legal ability.

Id.

149. MACCRATE REPORT, *supra* note 144, at 330-31.

150. James M. Peden, *The History of Law School Administration*, *in* 2 THE HISTORY OF LEGAL EDUCATION IN THE UNITED STATES, *supra* note 108, at 1105, 1125.

. . . compare the scores before and after [some particular] 'treatment' . . . [or] correlate scores with other [tests]."[151]

These methods are almost entirely unknown in the realm of the legal education community, whether due to a lack of knowledge, time, or priority. The failure to attempt these and other validity assurances perpetuate the ambiguities surrounding the examination process. Without comparison or a clear recitation of the objectives, teachers and students will continue the process as it has been administered over the past decades, still not knowing whether the evaluations are effective.

Even if these post-examination validity measures are used, the tests themselves may undermine validity. The tests may contain any of the following: "1. Unclear directions 2. Reading vocabulary [and] sentence structure [that is] too difficult 3. Ambiguity 4. Inadequate time limits 5. Inappropriate level of difficulty of [the test] items 6. Poorly constructed test items 7. [Test items that are] [i]nappropriate . . . for [the] outcomes being measured 8. Test [is] too short 9. Improper arrangement of items 10. Identifiable pattern of answers."[152]

In addition, validity issues spring up in the grading process as well. One problem immediately confronting a law professor about how to grade examinations is articulating objectives. Is the objective to rank students in their analytical reasoning abilities? Is it to include consideration of presentation factors such as spelling, handwriting, and neatness? On the other hand, is the examination more a measure of potential than of current ability?

3. *Reliability Defects In the Grading Process*

Even if law school exams test "the right stuff" and are valid, the grading process provides daunting reliability hurdles. Instructors must decide whether to evaluate students objectively, if possible, or holistically, based on the independent judgment of the grader.[153] Whichever method is chosen, reliability is an es-

---

151. *Measurement for Teachers: Module Three, Validity*, *at* http://ruby.fgcu.edu/courses/chewittg/10009/module3n.htm (last updated in 1999).

152. *Id.*

153. *See, e.g.*, Kissam, *supra* note 3, at 465 (stating that while the holistic method may provide a better measure of law student learning, there are great pressures to conform to objectively justified gradations of grading).

sential lynchpin of exams, and without it, a fair and accurate testing of valid objectives will not result. Thus, reliability is crucial to a fair test.

Yet, with instructors lacking knowledge of reliability principles and a dearth of institutional efforts to ensure reliability, law school exam results remain questionable. For example, a study involving seventeen law professors grading eighty law exams sought to determine if the professors shared a discernible basis in evaluating the examinations.[154] The results indicated there was no clear basis for assessing which responses were better than others.[155] Additional testing factors, such as writing style, time pressure, cultural background, and organizational abilities, interfere with measurements of performance,[156] and further skew the reliability of test results.[157]

Other reliability problems arise from the attempt to grade "objectively." The effort of assigning numerous gradations in a grade renders the analysis fragmented and piecemeal.[158] The grade distinctions are often so minute—a one point difference on a test with several hundred total points may distinguish one grade from the next—that the grading difference is not consistent with essay tests involving writing style, organization, and other hard-to-quantify variables.

The desire, even mandate, of many schools to impose grading curves exacerbates the reliability of grade distinctions. Whether called for or not, curves require instructors to distribute evaluations and make qualitative distinctions between

---

154. Stephen P. Klein & Frederick M. Hart, *Chance and Systematic Factors Affecting Essay Grades*, 5 J. EDUC. MEASUREMENT 197 (1968).

155. *See id.* at 199–201.

156. *See generally* W. Lawrence Church, *Law School Grading*, 27 WIS. L. REV. 825 (1991); Henderson, *supra* note 78; Roger C. Cramton, *The Current State of the Law School Curriculum*, 32 J. LEGAL EDUC. 321 (1982).

157. Redlich & Friedland, *supra* note 82, at 146.

158. Kissam, *supra* note 3, at 446:

> Sympathetic readings of essays are possible under the objective model of grading, but the piecemeal and fragmented nature of this method—as one searches for the many specific elements to a "right answer"—diminishes the likelihood of sympathetic interpretation. The objective model is more likely to produce unnatural interpretations, as professors read student essays in a negative state of mind in order to produce the many quantitative distinctions that are inherent in objective grading practices . . . .

Id. at 446–47.

papers that are sometimes quite similar in substance. Ironically, the imposition of a grading curve necessitates a more subjective, norm-referenced grading, in which professors judge students against the other students in the class, not against some objective, standardized measure.

The national standardized test-givers, such as the bodies who administer the Scholastic Aptitude Test (SAT) and Law School Admissions Test (LSAT) implicitly recognized the problems of reliability inherent in grading essay exams and switched to entirely objective multiple-choice type items.[159] These bodies recognized that grading essays is invariably a holistic enterprise that would not meet the requirements of standardized testing.

One initial problem is the lack of clarity in the professor's objectives and what the professor includes on an examination. Some professors like testing new material to see if students can use their analytical abilities to analogize from material covered in class. Other professors have multiple objectives that may conflict or be differentially represented on an examination. These objectives include: teaching legal analysis; covering substantive rules and principles; problem-solving; issue-spotting; exploring ethical considerations; using precedent and policy to argue persuasively; and teaching practice skills, among others.

4. *Other Defects in Reliability*

Even if timed essay items did test for important legal skills without significant distortion, these questions still suffer from considerable reliability problems. There is little consistency or predictability in grading, particularly with how to weigh writing ability, understanding of the rules, ability to apply the rules, and other important considerations. In essence, it is no wonder that overwhelming evidence shows that the essay test item is highly unreliable.[160]

Even if the weighting of factors problem has been resolved, a related reliability issue often arises—differential standard deviations for different parts of an exam. In effect, even if a pro-

159. Lawrence W. Ross Jr., *The Construction and Selection of Objective Business Law Examination Questions*, 17 AM. BUS. L.J. 547, 548–49 (1980).

160. Nickles, *supra* note 122, at 444.

fessor weighs all parts of an examination equally and could filter out extraneous factors, those parts with the greater distribution of scores will effectively carry more weight as to the outcome of the overall grade.

Furthermore, since students learn differently, a single type of examination question occurring in a single sitting will not be as reliable as utilizing different methods of evaluation with different types of questions. This is particularly true for essay exams, which some argue are graded as if they are objective exams.[161]

C. *The Undervaluation of Evaluation As a Teaching Tool*

If the evaluation process is overvalued as a measuring device, it is equally undervalued as a feedback tool to promote improvement in the skills of students. Its desuetude is due to tradition and misconceptions about law school learning.[162]

Many teachers perceive legal education as the delivery of information.[163] The teacher delivers the substantive material and the students receive and digest it, much like the delivery of the U.S. mail. The quantity and quality of material delivered,

---

161. Crane, *supra* note 3, at 787–88:

> To remedy the inherent unfairness of this tradition of testing the material so differently than the way it was taught, many law professors respond, almost intuitively, by grading the essay exams that they have administered as though they were not essay exams at all. Instead they grade the essay exams as though they were actually objective exams. They dissect the students' answers to the essay question into smaller sub-issues to which they assign a finite number of points on an answer key from which the student may earn credit, thus "objectifying" the essay examination.

Id. (footnote omitted).

162. *See* Kissam, *supra* note 3, at 436:

> The objectivism of law school examinations allows professors to limit their engagement with both the teaching and evaluation of their students. The marked discontinuities between classroom work and examination work and the use of quantitative methods to read and grade law school examinations are the primary means by which professors achieve this disengagement.

Id.

163. Instead, alternative conceptualizations of legal education exist. One alternative is to view legal education as a process system instead of as a delivery system, in which the goal is the improvement of student skills and performance, not a teasing out of potential. Within this framework, evaluations constitute a dynamic aid—instead of measuring skills—revealing to students what should to be done to improve. When used as a feedback tool, evaluation has a process orientation.

under this view, can be measured as a static amount at a particular point. This linear conceptualization does not comport with many studies of education, suggesting that students learn in different ways.[164] The studies indicate that the retention of information in short-term and long-term memory is not a "video camera" exercise, but something much more complicated, affected by circumstances such as student background, culture, learning styles, and other variables.[165] Given these variables, lecturing—the most common form of educational "mail delivery,"—is not suitable for maximum learning for the entire class.

The alternative process-oriented conceptualization, where the teacher's position is more of "coach" and "guide" to the learning process than "deliverer of information," plays only an indirect part of common law school practice. The so-called Socratic method of multiple questions and responses, without acknowledgment of "correct answers," is supposed to lead students to discover for themselves the appropriate framework of analysis for a particular problem and for the body of material as a whole.[166] While this may offer at least some feedback to students, it is generally so diffuse and non-responsive as to be of marginal use for specific agendas of improvement, certainly insufficient to highlight precise problems.

Further, it is assumed that performance in class is representative of examination performance. Students soon learn that the classroom may be a very poor indicator of examination performance, and that what a student perceives is his or her understanding of the subject matter may vary widely from perceptions about examination performance and the examination grade.

Even the feedback that may be gleaned from the final grade is likely skewed. Many schools have explicit or implicit curves, requiring numerous distinctions that may in fact lack true differences.[167] Other teachers may lump a wide disparity of pa-

---

164. HESS & FRIEDLAND, *supra* note 50, at 8–12.

165. *See* William Hathaway, *The Mind's DVD: What Happens When We Hit Replay*, ORLANDO SENTINEL, Sept. 11, 2002, at E1, *available at* 2002 WL 100139050.

166. Students are supposed to piece together, through inferences and connections, how legal analysis works, and how the course material is synthesized into overarching rules and principles.

167. *See* Kissam, *supra* note 3, at 449.

pers within a single grading category for the opposite reason—it is easier to give everyone an acceptable C+ or B, for example, than make more specific distinctions.

The piecemeal cutting and parsing resulting from the push toward objective grading further diminishes the feedback from exams. This pressure undermines qualitative evaluation of papers as a whole, forcing only discontinuous subsets to be the basis of the evaluation.[168] Thus, students do not receive a critique on their persuasive or advocacy writing ability in a holistic sense.[169]

1. *Educational Discontinuity*

The use of only one final test to assess a student's progress "prevents the test from providing any educational feedback."[170] In this regard, law school evaluations tend to emphasize critique, sometimes almost to the exclusion of observation.[171] Educational research indicates that educational development is fostered by multiple evaluations, not a single all-encompassing test at the conclusion of a course.[172] Multiple evaluations offer significant educational benefits. They "increase motivation, reduce test anxiety, increase facility with course material, and stimulate student efforts."[173] Conversely, a monistic approach has numerous educational deficiencies. By the time of the examination, and especially by the time grades are dispersed from

---

168. *See id.*at 447–50.

169. *See id.* at 446:

> [The holistic] method, unlike the objective grading style, invites professors to interpret and criticize student essays in the manner that we naturally employ to interpret and criticize works of art, social practices, and legal authorities—that is, first construing the work in a sympathetic way that provides the best possible interpretation of the work and then criticizing the work under explicit standards that pertain to a work of its kind.

See id. at 446.

170. Matthews, *supra* note 128, at 1104. "[T]he mechanism the school uses [to rank students] actually inhibits both learning and teaching." *Id.* at 1103–04.

171. An example of a critique-oriented evaluation would be a course with one final examination. These exams were graded from the "gut instinct" of the professor, who assigns A, B, C , D and F grades.

172. Nickles, *supra* note 122, at 461–62.

173. Henderson, *supra* note 78, at 412. ("Infrequent examination *is* an admission that testing is used only to assess the scholastic 'progress' of students, rather than to maximize the instructional possibilities."). *Id.*

those exams several weeks later,[174] the value of the examination as a course-related tool has diminished. Instead, there is a "disconnect" between the examination and the body of the course.

A second deficiency in the way examinations are used is the absence of the opportunity for reflection.[175] Generally, critical thinking is considered to include a step involving the considered relationship and meaning of raw data or facts to a particular problem. This step is glaringly absent from most final examinations in law school, except for seminars and clinical education. A time-pressured final examination rewards those students who can "shoot from the hip," and penalizes students who require additional time—even if only minutes—to formulate a cogent response. Further, there is a widespread, untested assumption that the final time-controlled essay examination captures and assesses the skills of "good lawyering." If analytical thinking is not the primary skill that is being taught in law school classes or if it is the primary skill but it is not reflected in test scores, alternative models of testing should be explored.

Lastly, there is another unstated assumption that high test scores—both during law school and on the bar exam—are indicative of a quality legal education. This assumption is questionable. If effective education is a product that relies on a number of factors, those factors must be included in the evaluation process. For example, frequency of evaluation, high expectations, cooperation between the school and the local bar and alumni, and clearly articulated goals all may play a large role in the quality of the education a school provides.[176]

---

174. For example, professors at DePaul Law School are given four to five weeks after an examination is given in which to grade the papers. Once grades are turned in to the appropriate official, they are distributed two weeks later. *See* Norman Redlich & Steve Friedland, *A Reply to Professor Jacobs: Right Answer, Wrong Question*, 41 DEPAUL L. REV. 183, 186 n.28 (1991).

175. "Learning theory suggests that reflection on the subject matter—and better yet, periodic assessment combined with reflection—provides essential feedback for the learning process." Henderson, *supra* note 78, at 412.

176. In a similar context, a member of Michigan State University studied the Detroit Public Schools and concluded that effective schooling was not correlative with test scores. Instead, factors that mattered to a good education included: "Strong instructional leadership[;] Clear and focused mission[;] Safe and orderly environment[;] A climate of high expectations[;] Time on task[;] Frequent monitoring of student progress[;] [and] A cooperative relationship between school and

2. *The Bar Exam and Further Educational Discontinuity*

The bar exam is not a direct component of the law school experience, but it often casts an extremely long shadow over that experience. From the very first year of school, students are solicited to sign up for bar review courses and urged to begin filling out their bar applications. Prior to completing the first year, students are confronted with what courses to choose for their second year of school. An immediate question is whether to take courses whose subject matter might appear on the bar examination. While student concern about the bar varies, for many students it is a brooding omnipresence and a not quite understandable rite of passage.

The discontinuity between testing and the substance of legal education continues with even greater force with the bar examination. This exam, now given in all fifty states, serves as a final academic barrier to law practice. The multiple-choice and essay questions purport to test the subtleties and nuances of the substantive rules and principles as well as legal analysis and problem-solving abilities. As one writer described the Multistate Bar Exam, "applicants spend six hours on one day trying to answer 200 tricky multiple-choice questions . . . ."[177]

As unfamiliar as such an exam is, faculty disavowment[178] of helping students to prepare for such an exam provides another illustration of competing interests between faculty and students. Many teachers profess to teach without regard to the bar examination and even proclaim their independence from the profession's gatekeeper exam.[179] Many view helping students to pass the bar examination as a capitulation of legal education to the students' immediate needs, as compared to more lofty ide-

---

home." L. Bracy Jr., Letter to the Editor, *Effective Education Doesn't Rest on Scores or Tests*, DETROIT FREE PRESS, Jan. 17, 1999, at 3H, *available at* http://www.freep.com.

177. *Questioning the Bar Exams*, TIME, Feb. 25, 1980, at 44.

178. Faculty disavowment of assistance often occurs because of the poor image of teaching to promote bar passage; such teaching is considered one of the baser aspirations of legal education, particularly when it is compared to objectives such as teaching students to think like lawyers.

179. Ironically, the disinterested institutional approach to evaluation has helped to cement the pervasive testing orthodoxy, which made student assessment completely dependent on the monistic, "all-or-nothing" final examination at the conclusion of a course. For courses that stretched an entire year, there was generally only one examination a year.

als, such as teaching students to "think like lawyers." For students, the bar exam lurks in the shadows, affecting course selection, class satisfaction, and residual anxiety levels. This dichotomous approach further contributes to the "disconnected" nature of examinations for students.

The disenchantment of most bar exam test-takers extends well beyond the requirement of additional testing, to the further discontinuity of the examination and what they were learning in law school and will learn in legal practice. As one commentator declared, "The bar exam as it exists today does nothing to further a student's education or thinking ability."[180] The bar exam does not provide testing for the wide variety of skills an attorney must use in practice, and that which it does test occurs in an extremely artificial environment using a limited number of test forms—primarily multiple-choice questions.

Significantly, the bar examination is not focused on the individual test-takers, but on the "mandatory minimum" for competency in the attorney marketplace. Thus, the bar examiners are not concerned about what higher scores in fact mean, and the test-takers take away no real understanding about what those scores reflect for the future—other than a sigh of relief associated with passing the exam.

3. *Teaching to the Whole Class*[181]

To whom is the teacher teaching? This question becomes relevant if it is recognized that students learn differently. If a teacher lectures, some students will learn more than others. If a teacher uses role-plays, small groups, or writing assignments, there may be different groups of students that reap greater benefits.

A second question is perhaps even more important—who is learning from the teaching? Which students are benefiting the

180. Hansen, *supra* note 109, at 1231.

181. This is the question posed in a terrific article and video presentation by Professors Gerry Hess of Gonzaga University School of Law, Paula Lustbader of Seattle University School of Law, and Lori Zimet of University of California at Hastings School of Law. *See Teach to the Whole Class: Barriers and Pathways to Learning, at* http://law.gonzaga.edu/ilst/PubsResources/Kit.htm (last modified Nov. 16, 2001). Their video presentation and accompanying workbook was produced in part by a grant from the Institute for Law School Teaching, located at Gonzaga University School of Law.

most—if at all—from the teaching methods used, and how can progress be discerned? Discernment is a concern if professors do not regularly dispense feedback.

Further, examinations omit feedback about a student's overall effort or progress in skill inculcation. This is somewhat ironic in an era of "rights" and due process—the law usually provides an explanation for legal decisions of all forms, with the exception of exams. "Grading due process"[182] has not taken on enough momentum to alter the cloak of secrecy that usually surrounds the grading process.[183]

These questions are simply not considered important in the current evaluation framework, given the assumption that teachers teach and the students, as a singular and unified entity, learn. The measure of this learning is the final examination.

4. *The Costs of Additional Teaching Tools*

Even if the evaluation and other teaching tools could promote the learning process, their costs arguably counsel against their use. Many professors claim that using evaluations as a teaching tool poses both time and competency problems. Evaluation during class time competes with course coverage of material. Time spent on evaluation could be put to good use covering additional cases or analyzing the existing cases in a more comprehensive manner.

IV. A PRESCRIPTION FOR IMPROVING LAW STUDENT EVALUATION

A. *In a Perfect World*

Imagine the following approach to evaluation in a newly formed law school: Student progress is measured by a variety of evaluation tools, such as oral quizzes, written interviews, short papers, and performance exercises, all on a periodic basis, and all geared toward articulated and clearly-stated institutional goals, such as a recognition that students learn differently and

182. Kissam, *supra* note 3, at 448 ("A third internal cause [of a shift in grading toward objective methods] is the contemporary law students' demand for 'grading due process,' or at least some explanation of their law school grades.").

183. *See id.* at 444–52.

for different reasons. The faculty gathers regularly to discuss the form and content of evaluations and work together to maximize their value, specifically forsaking the assumption that their own "tried and true" evaluations are accurate and valid measurements.[184] Many evaluation exercises are not graded and occur during class, serving as feedback and as the progenitor of communicated expectations, rather than as a creator of hierarchy and rank. In fact, evaluations are woven into the course pedagogy, are not forced to form an artificial curve, and are approved and administered in part by the students themselves.[185]

This prescriptive vision depends on using evaluation for two primary purposes: as a measuring device and as a learning tool. When used as a measuring device, it is best administered with greater institutional and personal scrutiny, recognizing that valid and reliable scores are difficult to achieve. When used as a learning tool, providing feedback and information, it can be of great assistance to the in-course improvement and advancement of students.

B. *Increasing Institutional Oversight*

Like other powerful institutional devices, the law schools should accord the evaluation process institutional oversight. In addition to scrapping the marginalized monistic system, closer scrutiny of the use of evaluation both as a measuring device and as a teaching tool is warranted. The value of law school exams as a measuring device is overstated, and its value as a teaching tool is understated. Instead, evaluation should be seen as serving differing and important purposes, and accorded the attention and oversight it deserves.

While institutions are aware of some of the drawbacks and deficiencies of the process, they pay little attention to improving

---

184. As one observer of the role of testing generally has noted: "After decades of agonizing over the fairness of S.A.T. scores, the differences between male and female mathematical skills and the gaps in I.Q. between various races and ethnic groups, the notion of intelligence and how to measure it remains more political than scientific, and as maddeningly elusive as ever." George Johnson, *Clueless; Tests Show Nobody's Smart About Intelligence*, N.Y. TIMES, Mar. 1, 1998, §4, at 1.

185. Suppose those same students were interested in learning about a wide array of material, even material lying outside the scope of an examination.

the system.[186] Part of the reason for the invisibility of the evaluation process is an acceptance of what initiated professorial success in the first place—the essay final examination.[187] In addition, with considerable demands on faculty to publish, teach, and participate in service-related activities, the evaluation process occupies a low priority.[188]

1. *Training and Guidance*

It would not be difficult for law schools to offer some examples or tips on preparing exams from an expert in psychometrics—or even veteran law professors. The examples would indicate the variety of questions, formulations, and constraints professors could use with their evaluations. Tips could include short lists of "do's and don'ts."[189] Outside experts could suggest different exam formats or question types. For example, one professor contacted a psychometrician who offers a guide to creating multiple-choice questions.[190]

Along the same lines as proposed training, schools could offer informal guidance on the subject of evaluation. A school could create a web site, a library of evaluation resources—such as Professor Michael Josephson's two volume set on examination and testing in law schools[191]—or simply promote informal discussions about evaluation on an institutional level, particularly involving veteran teachers.[192]

---

186. *See* Henderson, *supra* note 78, at 400–01.

187. "Most law professors are likely to perpetuate the system because they experienced the same system and prospered under it." *Id.* at 405.

188. *Id.* at 405 (labeling the perpetuation of the ranking system as institutional inertia).

189. For example, teachers could be informed about widely accepted "do's and don'ts," such as grade on-time, test what was covered in the course, provide clear and complete directions, and grade with a clearly determined process. Don'ts include: don't ruin the school's curve, if any, don't change grades without justification, don't create an insuperable test, even for the top test-takers, don't have typos, and don't hand in the grades late.

190. Crane, *supra* note 3, at 794 n.45.

191. 2 Michael Josephson, Learning & Evaluation In Law School (January 1984) (unpublished manuscript submitted to the Assoc. of American Law Schools Annual Meeting).

192. Tips might include adopting something like Professor Karen Gross' "Three Sentence Summary," for which she asks first semester, first-year students to summarize the main points in a class in three sentences. Karen Gross, *Three Sentence Summary*, *in* TECHNIQUES FOR TEACHING LAW, *supra* note 50, at 275.

C. *Improving Evaluation as a Measuring Device*

"[There has been a] turn in social theory toward the view that most of our cherished categories are not absolutes but conventions that change across time and place . . . ."[193] —Orlando Patterson

Evaluation in law school does not have to proceed in the way it is currently designed. Instead, it can be readily reinvented, despite competing demands for an instructor's time.[194] Some suggestions on how to accomplish this are discussed in the following sections.

1. *Adopt Multiple Evaluations With Varying Item Types*

Evaluation in law school would better serve its function as a measuring device[195] if it were utilized early and often, including during the main body of a course. A sufficiently large number of tests serve to promote validity[196] by decreasing the likelihood that a single, end-of-the-course exam excluded or distorted appropriate test topics. Multiple evaluations also minimize the likelihood of students suffering from an "off-day" during testing. The multiple evaluations do not have to be equivalent to several whole final examinations, but can constitute shorter tests or quizzes.

Since students learn differently, fairness dictates that instructors should include different types of test items in the evaluation process.[197] Different types of questions will allow all

193. Orlando Patterson, *America's Worst Idea*, N.Y. TIMES BOOK REV., Oct. 21, 2000, at 15 (reviewing SCOTT L. MALCOMSON, ONE DROP OF BLOOD (2000).

194. Competing demands include teaching, scholarship, and committee work. If the true importance of evaluation is recognized, however, teachers might make room for more serious pursuits of evaluation, creation and implementation. Further, teachers might be more aware of how to circumnavigate obstacles to anonymous and fair exams.

195. Students are measured for purposes of distinguishing among classmates in course performance, for prospective jobs, and simply for intrinsic merit—a good grade is the equivalent to the proverbial "job well-done."

196. *Cf.* Crane, *supra* note 3, at 794 (stating that "[t]he number of questions asked is important to the validity of a test because of the need for an adequate number of samples of each student's knowledge of the material being tested.").

197. While the traditional essay final examination serves a useful purpose and has been roundly defended on many occasions, multiple evaluations of varying types with a sustaining number of questions would promote the accuracy of the measuring process. *See, e.g.*, Redlich & Friedland, *supra* note 174 (proposing that

segments of the class to be on equal footing—to have some question-types that allow them to display their test-taking strengths. It follows from this premise that an examination with a variety of question-types in varying formats will promote the highest level of content validity. Those who prefer essays and those who prefer multiple-choice questions, for example, will all be able to exhibit their strengths.

In addition to multiple evaluations of varying types, it is important to ensure validity by including a sufficient number of questions overall. If there is only one question, such as a single essay focusing on only a small part of a course, or even a larger essay which has a "gatekeeper" issue that a student must spot and deal with properly before reaching the rest of the issues, the evaluation may not be representative of a student's mental ability, knowledge, or potential lawyering skills. In turn, this will undermine the validity of the evaluation enterprise.

The effort to ensure validity in testing underscores the potential invalidity of the traditional complete essay final examination given in most first-year and core law school courses. Instructors write these essays in a condensed form containing dependent, or even "gatekeeper" issues, saddled with a general and ambiguous call of the question such as "discuss the issues," and answered by students in a single, time-pressured sitting. These limitations diminish the likelihood that the test will reveal a student's true mental abilities or other skills.

While the validity of multiple evaluations of varying formats improves with frequency and variety, content validity still depends on the existence of well-defined tasks within the evaluations purporting to measure appropriate skills or mental processes. If the test does not evaluate appropriately defined tasks in a clear and unimpeded manner—impeded by, for example, time problems, ambiguities, or a failure of the test to be representative of the course—then the test is still invalid.

Other test item forms can include short-answers and multiple-choice. Many teachers are unfamiliar with the creation and implementation of multiple-choice questions, and thus shy away from them as a practical matter. Still other teachers do

---

objective test questions, such as multiple-choice, should be used as a supplement to the traditional final essay examination).

not believe that objective questions can test thinking ability, at least in the same way as essays. While such questions may be more difficult to create—if only because of their initial novelty—and may indeed test thinking in different ways than an essay, such questions have been shown to be a successful test of law school students.

One professor consulted a psychometrician for his suggestions on creating multiple-choice questions. The expert suggested the following:

> (1) "[e]ach item should be based on a single, clearly defined concept rather than on an obscure or unimportant detail"; (2) "[u]se unambiguous wording. Be precise. Students' performance should be related to their knowledge of the subject, not their ability to decipher the meaning of the question"; (3) "[m]inimize reading time. Unless you must use lengthy items to define the problems, as when complex problem solving skills are involved, long reading passages are likely to reduce reliability"; (4) "[I]t is acceptable to vary the number of alternatives on the items. There is no psychometric advantage to having a uniform number . . ." (5) "[d]o not [include] double negatives . . ." and (6) "[a]void systematic patterns for correct responses."[198]

2. *Measuring What? Articulating Specific Test and Course Goals*

Evaluations should test for specific, well-articulated objectives. It is not enough to say that an essay examination tests students' abilities to "think like lawyers," unless it is extremely clear how competent lawyers think. Articulated objectives, particularly those that are subject to measurement, are crucial to ensuring the accuracy and validity of test instruments.

There are numerous goals of legal education, most of them loosely connected to lawyering and the practice of law. Several of the most salient objectives, however, can be defined by using the following triumvirate: substantive knowledge, background education, and skills.[199] These objectives are achieved through

198. Crane, *supra* note 3, at 794 n.45 (quoting a memorandum from Dr. Julian J. Szucko, Ph.D., Applied Psychometric Services, to Professor Linda Crane, The John Marshall Law School (June 26, 1999) (first, second, third, fourth, fifth, and seventh alterations in original) (on file with New England Law Review)).

199. The mix of substantive coverage and skill training has been widely discussed in recent years, particularly in light of the Report of the Task Force on Law

substantive coverage and skills training.[200] The primary skill tested is analytical reasoning. This type of thinking differs from experiential thinking, which relies on a person's existing experience, or creative thinking, which does not rigorously involve deductions about specific circumstances after applying general principles.

Analytical reasoning in law school often involves the application of general principles to specific sets of facts to discover the relationship between rules and circumstances. Many view the thinking process, particularly the analytical type,[201] as a multi-step process, including components of data gathering and reflection.[202] Appellate case reports are illustrations of this method. This type of thinking is considered to be a derivative of the scientific method.[203]

The evaluation process is intended to assess objectives as accurately[204] and efficiently as possible.[205] To promote appropriate objectives, teachers should create tests utilizing a construction validity design. One such design follows:

Construction Validity Checklist for Law School Exams:[206]

1. What are the precise tasks to be measured?
2. What mental processes are required by each test item?
3. Is there any comparison available for reviewing the outcomes, such as between different groups, other tests, or overtime?

Schools and the Profession, more commonly referred to as the MacCrate Report. *See, e.g.*, John J. Costonis, *The MacCrate Report: Of Loaves, Fishes and the Future of American Legal Education*, 43 J. LEGAL EDUC. 157, 171–80 (1993).

200. Keyes & Whincop, *supra* note 6, at 12 n.42.

201. Other types of thinking exist, including experiential, based on interactions and observations in the world, and creative, based on use and participation in the creative arts.

202. Reflection often takes time for pondering and contemplating facts.

203. *See* Sheppard, *supra* note 44, at 25–26.

204. Whether the law school evaluation process accomplishes its mission is largely unknown.

205. *See, e.g.*, Crane, *supra* note 3, at 792 ("The primary objective of testing is to reach a valid and reliable determination of a student's proficiency in the assigned material."). *Cf.* Johnson, *supra* note 184, at 1 ("[T]he notion of intelligence and how to measure it remains more political than scientific, and as maddeningly elusive as ever.").

206. *See Measurement for Teachers: Module Three, Validity*, *at* http://ruby.fgcu.edu/courses/chewittg/10009/module3n.htm (last updated in 1999).

4. Does the test itself minimize validity obstacles by providing:
   a. Clear directions?
   b. Appropriate time limits?
   c. Understandable hypotheticals (for students of all cultures and backgrounds)?
   d. Appropriate levels of difficulty for the test items?
   e. Competent construction of the test items?
   f. Appropriate test-taking conditions?[207]

a. *Measuring Critical Thinking*

Assuming that school officials intend to use a large part of the first year of law school and beyond to teach methods of critical thought, how can the evaluation process measure a student's ability to think, analyze, and evaluate?

Critical thinking has been defined as "reasonable and reflective thinking that is focused upon deciding what to believe or do."[208] As this definition recognizes, critical thinking requires or involves time for reflection and deliberation. The process of thinking, therefore, should involve a period dedicated to sifting, culling, and organizing data or input prior to the drawing of inferences or deductions. Following that period, inferences and deductions are drawn and redrawn. It may take many inferences and additional raw data for conclusions to be reached. These conclusions may be termed insights or simply command action.

In law school, legal writing can be seen as a vessel in which critical thinking occurs. Yet, most good legal writing requires rewriting, suggesting that first reactions may prove to be only a temporary and relatively inaccurate reflection of a person's ability to solve legal problems. On the other hand, careful editing—a form of critical thinking process—often occurs on a regular basis with serious legal writing and may be a more important skill than the ability to write a coherent first draft.

---

207. *See id.*

208. Stephen P. Norris, *Can We Test Validly for Critical Thinking?*, 18 EDUC. RESEARCHER 21 (1989) (quoting STEPHEN P. NORRIS & ROBERT H. ENNIS, EVALUATING CRITICAL THINKING 1 (1989)).

Thinking can be, and often is, measured indirectly. A person's conduct may reference a person's thinking. Similarly, a person's writing can be a reflection of thinking ability.

1. *Domain Specific Thinking*

A trend in the current literature on critical thinking suggests that the quality of a person's thinking depends on the particular context or domain in which the thinking occurs.[209] In essence, the quality of the thinking depends on the circumstances or locus in which it is measured.[210] Consequently, it is not enough to speak of "good thinkers" or "bad thinkers," but rather of good and bad thinkers in particular fields or contexts.[211]

2. *Orders of Thinking*

Just as a person's thinking is affected by the particular context in which it occurs, there are different levels or orders of thinking as well. Professor Benjamin Bloom, in a seminal work, posited that types of thinking can be identified based on levels of complexity.[212] These orders of thinking, in ascending order, include knowledge, comprehension, application, analysis, synthesis, and evaluation.[213]

3. *Orders of Thinking in Law School*

A law school can adapt Professor Bloom's pyramid of learning to its educational process. For example, it is widely acknowledged that mere memorization of rules and principles is a lower order of thinking, and consequently, less valuable than the application of rules and principles. One commentator has suggested a pyramid for legal education, with synthesis the

209. *See* Paul T. Wangerin, *"Alternative" Grading In Large Section Law School Classes*, 6 FLA. J. L. & PUB. POL'Y 53, 61 n.13 (1993).

210. *Id.*

211. "[G]ood thinkers in particular fields often possess more information than poor thinkers in that same field." *Id.*

212. TAXONOMY OF EDUCATIONAL OBJECTIVES 15–18 (Benjamin S. Bloom ed., 1956).

213. *Id.* at 18. *See also* Josephson, *supra* note 41, at 53 (discussing Bloom's taxonomy of cognitive learning).

highest order of cognition, followed by judgment, problem-solving, issue-spotting, understanding, and knowledge.[214]

4. *Measuring Learning, Not Teaching*

Teachers concentrate too often on what they are teaching and not what students are learning. It is broadly assumed that teaching and learning constitute an identity, and therefore, what students actually understand is a useless measure. Yet, the literature shows the fallacy of this assumption, and that students learn and respond differently to teaching. Thus, it may be helpful to work backwards and determine if students are learning what it is they are supposed to be learning.[215]

One commentator, for example, has suggested that legal education revolves around four central questions.[216] These questions are the core of the Socratic dialogue favored by many law professors. The questions are:

1: *Why* are the rules/principles of law the way they are?
2: *What* are the rules/principles of law?
3: *How* do you solve problems using the rules/principles of law?
4: *What if* the facts were changed in these problems?[217]

If these questions supply the salient perspectives for case analysis, professors should consciously incorporate them into the formulation of law school evaluation. They could ask the questions directly or indirectly in varying contexts.

b. *Measuring Competencies*

It has long been assumed that lawyering revolves around a central competency—critical thinking. Yet, the trend has been to view lawyering as comprised of many different competencies. One commentator, Professor Greg Munro from the University of

214. Josephson, *supra* note 41, at 58.

215. Two commentators illustrate this approach through the title of their handout to other teachers. Ruta Stropus & Charlotte Taylor, Helping Students See the Forest for the Trees: Ensuring a Solid foundation for the Exam (Oct. 2000) (handout at the SALT Conference Presentation 2000 and on file with Pace Law Review).

216. This analytical construct was suggested by Professor Grayford Gray of the University of Tennessee School of Law.

217. Grayford Gray, Remarks at the A.A.L.S. conference on New Ideas For Experienced Law Teachers (June 1995).

Montana, suggests that multiple competencies exist, and consequently, multiple assessments are appropriate.[218]

Competencies can be divided into three primary categories: knowledge, problem-solving, and critique. Knowledge is comprised of recall and understanding. Recall means to remember the precise rules, principles, and information upon which the law is premised. Problem-solving is essentially the application of legal rules and principles to particular facts and circumstances. Critique, the most difficult competency, asks students to extrapolate from a particular case or circumstance a generalized rule or principle. These competencies can be charted in a test matrix, organizing the exam around these sub-parts.[219]

c. *Measuring Skills—Implementing the MacCrate Report*

The study that led to the MacCrate Report suggested that law schools should be doing a better job in teaching lawyering skills other than analytical thinking.[220] These skills included negotiation, interviewing, client counseling, and oral and written advocacy.[221] Including such skills on law school evaluations will signify their importance, particularly if evaluations in the first year of school—evaluations of primacy—contain components assessing these skills. There could be a performance component as a part of such exams, similar to the trend with the bar examination, and short questions involving negotiation or client-counseling skills.

d. *Measuring Accurately*

For objective questions in particular, one way to promote fairness and safeguard the accuracy of the measurement is to permit students the opportunity to explain their answers. The opportunity to explain an answer is extremely useful when students confront perceived ambiguities, whether on essay or objective test items. Examination instructions should recommend

218. *See* MUNRO, *supra* note 76, at 111–127.

219. *See* CENTER FOR TEACHING AND LEARNING, FOR YOUR CONSIDERATION #8, IMPROVING MULTIPLE CHOICE QUESTIONS (November 1990), *at* http://ctl.unc.edu/fyc8.html (November 1990).

220. *See* MACCRATE REPORT, *supra* note 144, at 4–7.

221. *See id.* at 138–40.

explanations under such circumstances and offer a guide on how to do so. Explanations also are helpful for objective questions, especially multiple-choice test items, permitting a better balance between the subjectivity of essays and the objectivity of selected response type questions. This form of question conversion has been called "answer justification," which essentially permits students to turn difficult multiple-choice questions into short-answer problems at their own discretion. One study of this technique found that "[o]ver 90 [percent] of both users and nonusers like having [answer justification] available to them and recommend its use in other classes."[222]

D. *Using Evaluation as a Learning Strategy*

"[T]he horizon will move whenever we do, that there will be a context beyond every context, a perspective that transcends any previous perspective."[223] —Richard Rorty (paraphrasing Nietzsche)

Teaching and learning in law school primarily revolve around a narrowly crafted learning style called the "Socratic Method." This methodology involves a teacher asking students a series of questions about an issue. The students are expected to learn about the law—and legal analysis in particular—from the exchanges, and not so much from direct responses by the instructor.

Yet other learning methodologies exist and some are more effective than the others depending on the particular student. The context in which the learning process occurs further influences the existence of differing learning methods. The particular professor teaching the class, the physical plant, the time of day, and the duration are just some of the variables that potentially influence learning. All of these variables lead to the conclusion that different learning strategies, whether adopted as a

222. David K. Dodd and Linda Leal, *Answer Justification: Removing the "Trick" from Multiple-Choice Questions*, 15 TEACHING OF PSYCHOL. 37, 37 (Feb. 1988).

223. Richard Rorty, *Freud KO's Plato*, N.Y. TIMES BOOK REV., Oct. 22, 2000, at 14 (reviewing JONATHON LEAR, HAPPINESS, DEATH, AND THE REMAINDER OF LIFE (2000)).

supplement or complement to the traditional Socratic Method, may indeed improve the classroom learning.

Many in other educational fora assume that feedback is an essential ingredient for advancement. Parents provide their children with daily feedback as part of the socialization process, and in sports and music, teachers provide regular observation and feedback for students to use as improvement. When a music teacher gives a lesson, the teacher observes the student play and offers a two-part critique—an observation of what the student appears to be doing and then a critique on how to improve that particular conduct.

One such strategy involves an instructor's expectations concerning student performance. In 1968, Professor Robert Rosenthal studied the role of teacher expectations and concluded that "when teachers expect students to do well and show intellectual growth, they do; when teachers do not have such expectations, performance and growth are not so encouraged and may in fact be discouraged in a variety of ways."[224]

Expectations can be changed, for example, to include evaluation not just as a measuring device, but also as an effective learning strategy.[225] Some of the benefits of increasing the emphasis on the educational properties of assessment appear below.

1. *Evaluation as Notice to the Students: Are Students Learning What They Ought to Learn?*

Due process of law requires notice, or notice and a hearing, before the government can deprive a person of a property or liberty interest.[226] Similarly, prior to evaluating law students on their knowledge, skill, and problem-solving proficiency, students should be fully informed about what they should be learning. In addition, students should be informed about the content of the impending evaluations if they are different—in either form or substance—from what the students should be learning.

---

224. James Rhem, *Pygmalion in the Classroom*, THE NAT'L TEACHING & LEARNING FORUM (James Rhem & Associates, Inc., Madison, W.I.), Feb. 1999, at 1 (discussing ROBERT ROSENTHAL & LENORE JACOBSON, PYGMALION IN THE CLASSROOM: TEACHER EXPECTATIONS AND PUPILS' INTELLECTUAL DEVELOPMENT (1992)).

225. *See* Barnes, *supra* note 14, at 178.

226. *See* U.S. CONST. amend. V.

The more importance and clarity attached to the notice to the students, the more likely the students will pay attention to such insights. With student distractions often common, an oblique or "soft" reference to the learning process often is insufficient to reach the whole class.

While a discussion about the general substance of the test often is a step in the right direction, it may not help all of the students to understand what it is exactly they should be doing to prepare for such an evaluation. A better way to meet an implied due process limitation is to offer simulated evaluations during the course. This allows the students to understand what it is they are supposed to be learning and enables them to make appropriate adjustments during the semester if they are not. Professor Gerry Hess offers students "minute papers" as a means of promoting feedback about what students are learning.[227] In these papers, Professor Hess asks students, "What is the most important thing you learned in class today?" and "What important question remains unanswered?" The results are then reported to the class.[228]

The more specificity an instructor uses in articulating the objectives, the better.[229] The instructor can articulate the objectives in the course syllabus, on the blackboard in class, in handouts, and through simulated evaluations. Students would learn, apart from the substantive material, what a knowledgeable person would know from the course, and what technical mastery must occur to succeed.[230]

---

227. Gerald F. Hess, *Minute Papers*, *in* TECHNIQUES FOR TEACHING LAW, *supra* note 50, at 269.

228. *Id.*

229. Sample objectives in Constitutional Law: (1) Learn about constitutional history; (2) Learn the basic doctrines and how they apply; (3) Learn about constitutional arguments; (4) Learn how constitutional law is made; (5) Articulate and defend the preferred balance of power in government—courts versus the legislature versus the executive.

230. Some key questions might be: (1) What would a knowledgeable person know? (2) What distinctions could such a person make? An example of what a knowledgeable person taking an introductory class in property law would know: (1) An executory interest cuts short the preceding interest and a remainder does not; (2) The right to travel under the Equal Protection Clause is often triggered by a law prescribing a durational residency requirement; (3) Impeachment by a crime of dishonesty or false statement is permissible regardless of whether the crime is a misdemeanor or felony.

2. *Evaluation as Notice to the Instructor*

The feedback provided by in-class evaluation can benefit instructors as well as students. Professors often have little idea as to how well students are learning during the semester, especially which techniques are working, which are not, and what students find most helpful. Student questions during the course of discussion may provide some insight, but such indirect anecdotal evidence is quite inconclusive. For novice and veteran teachers alike, patterns and habits soon develop in law school teaching. These patterns and habits may result from consciously created goals for each course, from the emulation of one or more role models, or from other sources. While teachers may receive feedback on their habits through peer review, student evaluations, or other sources, it is rare that teachers receive similar feedback on their examinations. Posing these questions prompts self-assessment of the evaluation process, and may lead to novel and significant approaches to the subject of student assessment.

The instructor can also obtain direct feedback from the students about what they most or least understand.[231] Alternatively, a test can provide useful, if not identical, data for the professor.

a. *Evaluation of the Evaluation*

Students are asked to complete written anonymous evaluations about a course at or near the end of its classroom component.[232] This time frame means that the evaluation process is routinely omitted from scrutiny. While students can and often do comment on a course, from its pedagogy, to its materials, to even the professor's moods and clothing, there is no opportunity for students to evaluate the evaluation. Creating a post-exam

231. Professor Grayford Gray of the University of Tennessee Law School asks his students to respond in writing to the following questions: "What about this material is most (1) engaging? (2) alienating? (3) surprising? (4) confusing?" Grayford Gray, Remarks at the A.A.L.S. Conference on New Ideas For Experienced Law Teachers (June 1995).

232. Many instructors, such as Gregory Munro, desire a more effective evaluation of the course and the instructor. *See* Gregory Munro, *More Effective Student Evaluation of the Course and Instructor*, *in* TECHNIQUES FOR TEACHING LAW, *supra* note 50, at 281–83 (proposing use of the small-group instructional diagnosis system (SGID), created at the University of Washington for undergraduate students).

evaluation, either a special one or waiting until the examination has been completed, would provide additional and helpful feedback about the evaluation process.

Further, an alumni course-evaluation can be given, assessing the value of the education from a distance. Delayed feedback may be even more valuable regarding lawyering skills that are best tested in the actual practice of law.[233] This perspective could augment the feedback given to professors about their teaching. Further, it may allow the students to take a more distant and disinterested view of the course, without either the favorable or unfavorable emotion attached to being in the middle of a class.

Professor Daniel Gordon utilized this device after teaching 600 students the subject of civil procedure during a nine-year period.[234] His goals in teaching extend from mastery of knowledge, to using that knowledge in "the work place context."[235] Professor Gordon received ninety-five anonymous responses from a wide variety of students.[236] The questionnaire provided a great deal of useful information, particularly as it related to the development of lawyering skills.[237]

3. *Evaluation as a Feedback Tool For Students*

Unlike a conceptualization of learning as delivery and reception, legal education can be reconceived as improvement—either in knowledge level possessed, the application of that knowledge, or in a variety of skills. This conceptualization moves dramatically away from the emphasis on final exam grades after the first semester of law school as the presumptive declaration of a law student's talent and whether the students

233. This perspective was adopted by Professor Rebecca Cochran of the University of Dayton School of Law, who uses "student consultants," certain second-year students who have accepted summer clerkships with law firms, to provide retrospective feedback on their first-year legal research and writing skills course. *See* Rebecca Cochran, *Feedback to Teachers: Professional Skills Student Consultants*, *in* TECHNIQUES FOR TEACHING LAW, *supra* note 50, at 271–73.

234. Gordon, *supra* note 50.

235. *Id.* at 57.

236. *Id.* at 61–2.

237. Professor Gordon noted that: "The MacCrate Report sees lawyer development and the legal educational process as a continuum, which begins with an individual self-assessment and self-development process in the years before a student even enters law school." *Id.* at 62.

"have it"—the elusive quality of "thinking like a lawyer"—or not. With this new viewpoint, classroom assessment becomes more important, particularly if it can be directly linked to improvement. Assessment techniques that are practice instruments not only provide opportunities for self-evaluation, but promote proper techniques in analysis and other pertinent skills.

This new conceptualization of evaluation is based on the premise that evaluations are connected to the rest of the legal education process, and shows that an examination is not merely a measurement addendum. Instead, evaluation plays an integral role in the learning process. Using multiple examinations, some during the actual course, reconfigures the function of the evaluation process. No longer is it all too easy to view examinations as a separate and distinct component from the rest of a law school course.[238] Questions about the examination process now directly relate to student learning, precipitating inquiries such as what exactly does the teacher expect the students to be able to accomplish, do, and know by the end of the course. These questions invariably call into play the deep-rooted values, beliefs, and objectives of the course. The result is a bridge or linkage with the classroom component of the course.

If the examination process is integrated with the pedagogical aspects of the course, students may be better able to assess their own strengths and weaknesses, and work on them in the future. For example, students may want to improve but may not know how. Evaluation can show students how to improve. In other words, evaluation helps to create a template of model skills or abilities, so that the practice afforded by the evaluation helps students to "make it their practice." Further, additional evaluation opportunities provide repetition and enhance the opportunity for assimilating and understanding principles and points of law.

Evaluation can provide students with a better understanding of the course material, but just as important, it can offer them a snapshot of their level of knowledge, what they need to

238. After all, a traditional examination often occurs days, even weeks, after the last class has concluded. The grading process is also distant and distinct from the organic interactions in the classroom, and the teacher's role is very different than that of instructor.

focus on to improve, and even how to make adjustments during the semester. The use of evaluation can occur in-class in many different forms. It could be a short quiz, a non-graded writing exercise, or even a written answer to a typical "Socratic" question—asked of an entire class instead of a single individual.

When evaluation is used as a feedback tool, it can occur inside or outside of the classroom. All that is contemplated is that it occurs during the substantive component of the course. Some teachers use classroom assessment tools. For example, peer-editing, group writing, student problem creation, and self-grading are just a few of the in-class techniques.[239]

Peer-editing involves students reviewing the writing of their classmates. While the editors likely do not have the experienced eye of a professor, both the editor and writer can benefit from obtaining another perspective about the writing—how it is received, whether it makes sense, and whether it is clear.[240] This points out to students that the writer's intention is of little significance if the exercise of reading the writing yields a different meaning or understanding of the written words.

Group writing in class is another assessment technique that can be used to promote writing skills. Students can be asked to write down as little as the answer to a Socratic question typically devoted to a single individual. This short writing exercise accomplishes several things. First, it promotes involvement by the whole class in an active manner. It also allows students to compare their written answer against the oral

---

239. At the SALT Conference in New York City in October, 2000, Dean Ruta Stropus and Dean Charlotte Taylor suggested twelve different classroom assessment techniques in their handout, "Helping Students See the Forest for the Trees: Ensuring a Solid Foundation for the Exam" (October 2000). These techniques included: 1. Teach yourself (teach a subject to others, including creating a lesson plan); 2. Write your own (have students create their own hypotheticals);. 3. Self-grading (ask students to grade their own papers by using a grading sheet supplied by the instructor); 4. Peer grading (ask classmates to swap papers and evaluate each other's work); 8. Form/function chart (have students develop a chart of the form and function of substantive concepts); 9. Defining features matrix (have students create a chart dividing up the salient features of different legal principles); 10. Problem recognition (issue spotting practice); 11. One sentence summary (have students summarize concepts covered in class). *See* Stropus & Taylor, *supra* note 215.

240. *See generally* Kathleen Magone, *Peer Editing Benefits You And Your Students*, THE LAW TCHR, *available at* http://law.gonzaga.edu/ilst/newsletters/fall96/magone.htm (Fall 1996).

responses given in class.[241] This self-assessment technique also promotes writing during the semester, a good preparatory habit.

Teachers also can "mirror" common mistakes in an extra class, showing how those mistakes are made—and how they can be avoided. Students, while earnest in their attempts to improve, may not have any conscious recognition of their bad test-taking habits until it is shown to them through a road-map leading to improvement.[242]

Another tool involves students creating their own fact-based problems relating to the course material. The premise upon which this technique is based is that if students can create problems that raise specific course-related questions, then this indicates a better understanding of the substantive material. The creators of these problems can answer them—since the act of writing an answer will reveal whether the students' narrative abilities match their problem-solving skills—or other students can answer the problems.

4. *Learning In Context—"Teaching to the Whole Class"*[243]

Great teaching is only measured by whether there is great learning occurring in the class. It is often assumed that all students learn the same, and that the teaching will be effective if it is aimed at the mythical "average" student, whether it is the average top student or average student in the middle of the class.

---

241. Writing often reveals defects or errors in thinking, particularly when compared to perceptions based solely on mental impressions. Students sometimes are under the false impression that they know or understand material, only to have their written explications indicate to the contrary.

242. HESS & FRIEDLAND, *supra* note 50, at 286.

243. This title is the same one used by Professors Paula Lustbader, Laurie Zimet and Gerry Hess in a presentation and workshop given to law school instructors to promote the idea that teaching should take into account who the students are in the class. This incorporates the idea that students learn differently and their dissimilarities should be taken into account by professors. *See supra* note 181.

a. *Working Backwards—Teaching From the Ground Up*

To implement teaching to the whole class, it is preferable to adopt the perspective of the students rather than that of a teacher. Students as well as teaching professionals can answer the question of what will make students learn optimally. This approach "from the ground up" is consistent with the literature that suggests student evaluations are valid and students are aware of—and have accurate perceptions of—obstacles and advantages to good learning.[244]

V. CONCLUSION

The subject of law school evaluation is widely overlooked on both institutional and faculty levels. Closer scrutiny of the evaluation process reveals that evaluations are powerful, influential, and disproportionately pervasive components of legal education that are connected to almost every aspect of a student's law school experience. Indeed, what emerges is that evaluations are undervalued as a learning tool and overemphasized as a performance-measuring device. These errors can have far-reaching and improper consequences. To correct these errors, evaluations should be more carefully implemented as objective measuring devices, using the touchstones of validity, reliability, and efficiency (fairness) in constructing multiple evaluations of varying test item types. Further, evaluations should be more widely used as in-class pedagogical feedback tools, providing rigorous and active practice of mental and other skills as a path to improvement.

244. *See* Gerald F. Hess, *Student Involvement In Improving Law Teaching and Learning*, 63 UMKC L. REV. 343, 352 (1998).