
Linda R. Crane
John Marshall Law School

Follow this and additional works at: https://repository.law.uic.edu/facpubs

Part of the Legal Education Commons

Recommended Citation

https://repository.law.uic.edu/facpubs/161

This Article is brought to you for free and open access by UIC Law Open Access Repository. It has been accepted for inclusion in UIC Law Open Access Faculty Scholarship by an authorized administrator of UIC Law Open Access Repository. For more information, please contact repository@jmls.edu.
Grading Law School Examinations: Making a Case for Objective Exams to Cure What Ails “Objectified” Exams

Linda R. Crane*

I. INTRODUCTION

The traditional law school examination consists of a few essay questions created by adapting the facts of court cases, actual situations with which the professor is familiar, or even fictional situations from literature or popular culture. Under the standards established by the American Bar Association,¹ full-time law school instruction consists of two semesters (or three quarters) of fourteen to sixteen weeks each with classes meeting for two to four hours each week.² In the law school classroom, the professor presents the material to students through a series of teaching modules that are designed to facilitate the learning of dozens of points of law, policy, and skills. Later, students are tested by the professor in order to assess whether they have learned enough to be granted a law degree, to sit for the bar exam, and to practice law. The assessment of each student’s profi-

---

² See id. Standard 304 states in pertinent part:

(a) A law school shall require, as a condition for graduation, successful completion of a course of study in residence of not fewer than 1,120 class hours, ... extending over not fewer than three academic years for a full-time student or four academic years for a part-time student.

(b) An academic year shall consist of not fewer than 140 days on which classes are regularly scheduled in the law school, extending into not fewer than eight calendar months. Time for reading periods, examinations, or other activities may not be counted for this purpose.

Id.

785
ciency in each law school subject is generally made by comparing the performance on the exam by one student against the performance of all of the other students who took the same exam at the end of the same course of instruction.

Typically, the law professor will teach the entire course by leading a discussion about the material after students have prepared briefs of cases and short problems that convey information about the material — and the learning of it — in a piecemeal fashion. At the end of the semester, the professor has an expectation that the student will have mastered all of the material well enough to be able to understand the entire subject in a comprehensive way, even though that is not the way the material was taught. Nevertheless, that is the way the class material is most often tested and how student proficiency is assessed.3

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,”4 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.”5 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, and upon which their entire class standing and future careers rest.6 Questions require a student to evaluate law or policy in order to resolve a problem, questions of novelty, originality, and judgment issues that are involved in both constructing and grading the exam can lead to controversial outcomes.7 Although it is important for students to be able to demonstrate an ability to see the big picture and to answer comprehensive questions about the subject, it is far more difficult to do this when the material is tested in a way that so greatly differs from the way that it was taught. This creates a dichotomy that leads to charges of incoherence, arbitrariness and subjectivity.8

4. Id. (alteration in original).
5. Id. at 440 (alteration in original).
6. Although some law professors have begun to administer mid-term examinations, or have started using other methods in order to pace the advancement of the students during the semester before the final examination, most law school courses are still tested using the traditional law school model of giving one examination at the end of the semester. In some cases, where courses are taught over the course of two semesters, the final examination is given only at the end of the second semester.
7. See Kissam, supra note 3, at 442 (citing J.A. Coutts, Examinations for Law Degrees, 9 J. SOC’Y PUB. TCHR. L. 399 (1967)).
8. See Steve H. Nickles, Examining and Grading in American Law Schools, 30
Of course, subjective influences on the professor are always potentially harmful to students. These influences include many non-substantive distractions, such as the grader’s reaction to a student’s poor penmanship, style of writing, gender,\textsuperscript{9} choice of ink color, choice of line spacing, etc.\textsuperscript{10} Subjectivity in grading leads to inconsistent scores among students of similar proficiency. Inconsistent scores among test-takers produce unreliable and invalid test results.\textsuperscript{11}

To remedy the inherent unfairness of this tradition of testing the material

\textsuperscript{9}See id. at 445 n.109. Although most law school examinations are graded anonymously, many students believe that a professor who is so inclined may identify the gender of a test taker because of feminine or masculine handwriting traits.

\textsuperscript{10}See Douglas A. Henderson, \textit{Uncivil Procedure: Ranking Law Students Among Their Peers}, 27 U. Mich. J.L. 399, 409-10 (1994). The author also discredits the practice of ranking students because it “devalues human beings, distorts the legal system into a cultural compactor, and diverts scarce resources from truly legitimate educational goals.” Id. at 400.

\textsuperscript{11}See id. at 409-10; see also Stephen P. Klein & Frederick M. Hart, \textit{Chance and Systematic Factors Affecting Essay Grades}, 5 J. Educ. Measurement 197, 198-201 (1968) (reporting the results of a study where seventeen law professors graded eighty essay answers and could not identify the basis for determining whether some answers were better than others). Essay exams test a student’s ability to persuade and to communicate arguments to others, therefore success on essay exams partially depend upon the receptiveness to the argument by the readers themselves. See W. Lawrence Church, \textit{Forum: Law School Grading}, 1991 Wis. L. Rev. 825, 829 (1991).

An essay test is judgmentally scored, usually with a rating scale or a checklist. The ideal procedures for scoring are highly involved and laborious (Coffman, 1971). A number of biases exist in scoring essays. Chase (1979) established evidence for an expectation effect on essay test scores, as well as the influence of handwriting, a finding which Hughes, Keeling, and Tuck (1983) replicated. The relation of sentence length to test score and the influence of penmanship on the test score are well documented (Coffman, 1971). In a more recent view, Chase (1986) indicated several studies showing the existence of racial and gender bias in scoring essays. Such biases are very serious threats to the validity of interpretations and uses of essay test scores.”
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. The primary purpose of the "objectification" of the essay examination is to avoid the fear of subjectivity often experienced by law professors when grading traditional essay examination questions. Consequently, law professors routinely grade their students’ answers to essay examination questions as though the questions had actually been asked in a multiple choice or true/false format.

This "objectification" of the law school essay examination will generally result in a fairer assessment of a law student’s performance because it reduces the subjectivity in the grading process by increasing the number of objective elements upon which the grade is based. Law professors who use this approach to grading their essay examination questions should be congratulated for recognizing the problems inherent in grading essay examination questions. "Objectifying" essay examination questions and answers provides at least some quantifiable basis upon which to evaluate actual samples of student performance. This approach to grading student answers to essay questions greatly reduces the risk that unfairly subjective and/or ad hoc considerations will influence the final grade granted to any given student.

An "objectified" law school essay examination grade is fairer than one that is not, but it does not go far enough. An objective examination is fairest of all because it: (1) will test law students using a format that is closer to the law school class format that was probably used to teach the material being tested; and (2) will be more likely to result in a grade that is statistically valid as a result of being based upon a sufficiently large number of samples of a student’s knowledge of the material being tested.

II. WHAT ARE "OBJECTIFIED" EXAMS AND WHAT AILS THEM?

After teaching a course using cases and problems that require students to analyze issues and sub-issues, and after dissecting the material into single-issue sections, law professors will routinely base the student’s entire grade...
for the semester on the results of an examination that consists of comprehensive hypothetical essay questions that require the student to demonstrate an understanding of the course material in a manner that bears little, if any, resemblance to the manner in which it was taught. "Surprise" essay exam questions are often "different from any problems analyzed during the semester." 16 Surprise is also achieved by masking the issue amidst a plethora of facts. 17 Although surprise may be needed to test the students' ability to spot issues, combining the element of surprise with severe time constraints hinders the exam process and the students' ability to develop detailed answers. 18 The typical law school examination is administered to all students at the same time over a two to four hour period, depending usually on the course's number of credit hours. To the extent that the professor drafts the examination so that the student will need most of the time allotted in order to answer the questions to her satisfaction, speed becomes an additional element that will also affect, and possibly distort, the students' final grade. 19

It is almost axiomatic to say that law professors would teach for free, but require a salary to grade final exams. This statement is understood by all as expressing the general feeling most of us have about the task of grading. We do not like to grade final examinations. This means that, for many of us, the task is a chore that is made easier if a student's paper is easy to read. Students with papers that are easier to read, from the professor's subjective point of view, will often receive a friendlier response from the professor resulting in a better grade. This is especially true if the professor is grading the exam without relying on an answer key — choosing instead to assign a grade based on more intangible reactions to it.

Grading an essay examination answer without comparing it to an answer key is accomplished, apparently, by reading the essay answer and then deciding what grade the student ought to receive based on the general "impression" it leaves on the grader. Grading an essay answer without assigning points, either consciously or unconsciously, to those parts of the answer for which credit may be earned, requires the professor to grade the entire answer as an undivided whole through a largely ad hoc and subjective process. This process requires the professor to decide how well each student impresses the professor with his or her general command of the material. Unfortunately, this is a determination that relies heavily on how the professor "feels" about the answer after reading it in its entirety, without actually assigning specific point value to any of its sub-parts. Even worse, each student receives a grade that is tied to a rolling standard.

17. See id. at 454.
18. See id.
One of the most frequent topics of conversation among law professors is grading. The traditional law school examination consists of two to three essay questions. Essay questions are usually lengthy hypothetical fact patterns that end with an instruction to the student to do something. Typically, the student is asked to complete one of the following tasks: (1) to write a memo covering all of the relevant legal issues; (2) to write a letter to the law firm's senior partner advising her how to resolve the matter without alienating the firm's biggest client; or (3) to pretend to be a judge who is writing an opinion at the close of all of the relevant arguments. In order to follow the instructions, the student must usually identify all of the relevant issues (and sometimes also the irrelevant ones) raised by the facts and write a well-organized discussion that includes an analysis of all of the issues in light of the facts as they affect each of the parties.

I believe that I can say, without fear of contradiction, that one of the most common approaches taken by law professors when grading long, comprehensive, hypothetical essay examination questions involves the "objectification" of the answer. I have coined this term in order to describe the process of grading the answers to an essay examination question as though it was actually several objective examination questions.

The typical model answer to an essay question can be easily divided into component parts that allow the grader to assign a small number of points to each part that can be added together before being converted into a final grade. This "objectified" essay exam has more in common with the less traditional (for law schools) objective examination that uses multiple-choice and true/false questions than it has with the traditional law school essay exam—at least with regard to the way it is graded.

When "objectifying" the answer to an essay examination question, the professor pre-determines a total number of points that the student may earn in several areas of issue identification, sub-issue identification, statements of rules of law and policy, analysis of facts, and conclusions. The professor will use these points of information to create an answer key against which each student's performance will be scrupulously measured. This approach is a deliberate attempt to improve the process of grading the essay examination answer and to reduce the danger of allowing subjective determinations to influence grades out of sincere concerns for fairness, validity, reliability, and consistency.

Law professors express different reasons for "objectifying" the essay examination answer when grading it. Some claim that it makes the exams easier to grade, while other professors say that it facilitates the creation of an answer key that can then be used during student exam review sessions. However, it is clear that "objectifying" the essay examination answer when grading it seems fairer, at least to the extent that this process of "objectification" seems to reduce the overall amount of subjectivity in the grading. Most law professors are deeply concerned about the fairness of their exams and believe strongly that less subjectivity and more objectivity
in the law school examination grading process is a good thing.

When law professors take an objective approach to grading essay exams, they do so by creating an answer key that assigns various numbers of points that a student may earn by writing a perfect answer to the various components of the question. "Objectification" of the essay answer occurs through this process of breaking down the essay question, and the student's attempt at answering it, into component parts for which points may be earned according to a scale that is used to measure each student's performance. The impulse of so many law professors to "objectify" their students' answers to essay examination questions is a result of an innate sense of fairness and an awareness of the need to apply the same criteria as often as possible to the measure of each student's performance evaluation. This is especially important given our traditional law school model of ranking all students relative to their classmates based on their performance on a single test at the end of the course.

The "objectification" of grading essay examinations is also allows law professors to compensate for the inconsistency in the way they test the subjects and the way they teach the subjects. It allows professors to adjust for this uncomfortable inconsistency when grading the exams. "Objectifying" the grading of law school essay exams makes it more difficult for professors to render overly sympathetic readings of some students' answers given the existence of the answer key and correct answers.

A law professor's decision to grade an essay exam, at least in part, as though it were an objective examination, reflects an adjustment that is a good and worthy response to the need to address a problem of large consequence to our students: fairness in grading. However, "objectifying" the essay examination in the grading process does not go far enough. Instead, it is only a partial solution. A more complete solution to these grading problems is to refrain from testing so differently from the way we teach. If there is value in grading the essay examination as though it was an objective examination, then there is much more value in administering the examination in the format that conforms with the way it is graded: namely, as an objective examination. Rather than merely "objectifying" essay examination answers when grading them, law professors should draft the examination using objective questions. By using objective examinations, law professors can cure what ails "objectified" essay exams and thus increase their reliability, validity, and fairness.

20. The maximum number of points may or may not include extra-credit points on the exam and/or extra-credit for class participation.
21. See Nickles, supra note 8, at 420 n.19.
III. MAKING THE CASE FOR OBJECTIVE EXAMS: APPLIED MATHEMATICS AND PSYCHOMETRICS

The primary objective of testing is to reach a valid and reliable determination of a student's proficiency in the assigned material. "The concept of test validation refers to the appropriateness, meaningfulness, and usefulness of the specific inferences from the test scores, [and] [t]est validation is the process of accumulating evidence to support such inferences." Specialists in the areas of psychometrics and applied statistics analyze the technical soundness of testing materials and have developed several conceptual techniques and mathematical computations to aid in the prediction, description, and control of a given set of test scores. Experts in these fields agree that a test cannot be considered an accurate portrayal of a person's knowledge of a given set of information unless it has two essential characteristics: validity and reliability.

The sets of information that are to be tested are referred to as "domains" and, in the law school context, these refer not to entire subjects such as Property Law, Contracts or Torts; but, rather to topics within subjects, such as adverse possession or easements in Property Law, consideration or damages in Contracts, or proximate cause in Torts. A simplified definition of reliability is the test's ability to consistently measure a student's proficiency in a domain. A simplified definition of validity is the test's ability to measure or to tap into the particular domain that the test
purports to measure. 30 "Two important ways to increase reliability are to make tests longer and to improve the discriminating ability of the items appearing in the test. For comparable administration times, the essay test yields a lower reliability than the comparable multiple-choice version." 31

Reliability also looks at the degree to which the sample/domain of items "represents the universe from which they were drawn." 32 The "degree of reliability in the test that will be satisfactory depends on the purposes and circumstances of the test." 33 Tinkelman further notes:

A test designed to compare individuals requires a higher degree of reliability than a test designed to compare groups . . . . The minimum acceptable reliability depends on the seriousness of the decisions to be made about examinees . . . . In general, the reliability of a test or test part, as defined above, is a function of the item intercorrelations (or the item-test correlations) and the number of items. Within limits, the higher the average level of item intercorrelations in a test, the more reliable it will be. For any given level of intercorrelations, the longer the test, the more reliable it will be. 34

A test cannot be valid unless it is also reliable. 35 Put another way, an exam cannot test what it is intended to test if it does not first yield consistent results. This is as critical for law school examinations as it is for exams in other disciplines. In fact, as noted above, the greater the seriousness of the impact of a given exam, the more important it is for the examination to yield reliable results. 36 Very few exams have more emphasis or more serious consequences placed on them than law school final examinations — traditionally the only exam given for a grade that can affect the test-taker's entire professional life. The presence of competition among the test-takers also increases the importance of test reliability. 37 This too, is a feature of most law school examinations, especially when they are graded according to grading curves that are designed to compare the relative proficiency of all test-takers.

"Part of measurement theory thus concerns statistical relations between the actual test scores and the hypothetical scores if all items in the universe had been administered." 38 Statistics is a branch of applied mathematics.

---

31. HALADYNA, supra note 29, at 27 (citation omitted).
32. Tinkelman, supra note 19, at 71.
33. Id.
34. Id.
35. Telephone Interview with Dr. Julian J. Szucko, supra note 22.
36. See Tinkelman, supra note 19, at 71-72.
37. See id.
38. NUNNALLY & BERNSTEIN, supra note 25, at 10.
that uses mathematics and logic to analyze and to describe data. It is very general and does not specify what type of data is being analyzed; but simply applies mathematical formulas and logic when analyzing data. Psychometrics, in contrast to simple statistics, is a more specific application of mathematics to the design and analysis of research in the measurement (including psychological testing) of human characteristics. Many applied statisticians have developed an interest in psychometrics — hence the designation "psychometrician." Both disciplines are deeply concerned with testing reliability and the ability of a test to consistently measure a test-taker’s knowledge of a given domain.

In order for an objective examination to be statistically valid, several things must happen. One of the most important elements of any valid exam is that it must have a sufficiently large number of questions. The 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.

Assume that a typical law school essay examination consists of three long essay questions. Assume also that in an attempt to achieve greater fairness than mere anonymous grading will allow, each essay question is "objectified" by being divided, for grading purposes, into six sub-parts. The professor will base the grade on eighteen (three multiplied by six) samples of the student’s knowledge. A division of each of the three essay

39. Telephone Interview with Dr. Julian J. Szucko, supra note 22.
40. See id.
41. Psychometrics is a branch of psychology that is also referred to as psychological measurement. Telephone Interview with Dr. Julian J. Szucko, supra note 22.
42. Telephone Interview with Dr. Julian J. Szucko, supra note 22.
43. See id.
44. See id.
45. Telephone Interview with Dr. Julian J. Szucko, supra note 22. Dr. Julian J. Szucko has detailed guidelines to aid professors in writing multiple choice questions. Particular guidelines are: (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. Two plausible distractors are better than three implausible ones”; (5) “[d]o not double negatives. Following a negative stem with a negative alternative benefits only the students who are able to follow logical complexity of such items”; and (6) “[a]void systematic patterns for correct responses.” Memorandum from Dr. Julian J. Szucko, Ph.D., Applied Psychometric Services to Professor Linda Crane of The John Marshall Law School (June 26, 1999) (on file with New England Law Review).
questions into ten sub-parts will yield thirty samples of the students knowledge upon which to base the grade. If we assume that the exam consists of five short essay questions, and assume further, that each is divided into five to ten graded sub-parts the total number of samples would equal twenty-five to fifty. By writing an examination of the same material using objective questions, the law professor is able to base the final grade on more samples of the student’s actual knowledge of the material being tested.

Experts in applied statistics and psychometrics use computations such as the Spearman-Brown formula to describe the relationship among and between variables such as the overall reliability of the test, the internal reliability of the test (the consistency between individual test items in measuring a given domain), and the number of requisite items needed to sample a particular domain.

The Spearman-Brown formula allows psychometricians to “predict with considerable confidence the degree of reliability [that] would be obtained by using [varying numbers of] ... similar questions in an examination.” Generally to be considered reliable, a test must have an overall consistency of 80% or .8. Furthermore, the optimal level of consistency between the items measuring a single domain should be at least 50% or .5. Because this optimal level is very difficult to achieve, a test with good internal reliability yields items that correlate anywhere from .25 to .5.

Manipulating the Spearman-Brown formula informs us that a test with

\[ r_{nm} = \frac{nr_{11}}{1+(n-1)r_{11}} \]

A statistical device that is an example of a psychometric formula, it was originally known as “Brown’s Formula ... [but] was independently derived in its more generalized form by Spearman.” Id. at 245. Spearman-Brown Formula: \[ r^* = \frac{nr^2}{1+(n-1)r^2} \]

Tinkelman, supra note 19, at 46, 71. However, today the application of the formula remains unchanged.

46. See infra Table III and accompanying text.
47. Telephone Interview with Dr. Julian J. Szucko, supra note 22.
48. See Wood, supra note 15, at 245. Brown’s Formula: \[ r_{nm} = \frac{nr_{11}}{1+(n-1)r_{11}} \]
50. See id.
51. See id.
52. Wood, supra note 15, at 245.
53. See Tinkelman, supra note 19, at 72; Telephone Interview with Dr. Julian J. Szucko, supra note 22.
54. Telephone Interview with Dr. Julian J. Szucko, supra note 22. When one takes into account the extreme importance of the single law school exam that potentially affects a student’s earning capacity, his or her ability to feed and clothe his or her children and pay back any number of loans and debt acquired through the law school process, even the optimal reliability estimates seem to be an inadequate basis upon which to determine a student’s actual knowledge of a given domain, much less of an entire subject.
an overall consistency of .8 and an internal reliability of .3 must contain approximately forty samples of the student's knowledge, or questions, per domain. Of course, the typical law school course has several domains. Recall, for example, that in the subject of Property Law, adverse possession and future interests are considered two separate domains. A test of adverse possession alone would need to be comprised of at least forty questions. A valid examination that tests the entire subject of Property Law would need to consist of the forty adverse possession questions and forty questions from the future interests domain, as well as forty questions from every other discreet subset (domain) of Property Law being tested in order to measure consistently each domain at a rate of 80% with the items in the domain in agreement at a rate of 30%.

In situations where test results are used for important individual diagnosis, as is commonly the case of law school examinations, one may wish to vary the Spearman-Brown formula to assure at least a minimum level of reliability assuming a given level of item-test correlation: \[ r_{it} \] One would probably want to use .80 or .90 for \( r_{tt} \) (the reliability of the test or test part). Generally, .80 is the lowest reasonable level \( r_{tt} \) that one can use for a class that is given quizzes and multiple exams.

Given the importance and weight of the single law school final examination, law professors should consider striving for a reliability of at least .90 for the \( r_{tt} \) and solving for the unknown variable \( n \) (the number of items/questions necessary to secure an \( r_{tt} \) of .90). Dr. Julian Szucko, the Director of Test Development for the University of Illinois at Chicago recommends a starting \( r_{it} \) of .25 when solving the equation and advises that the best way to calculate this variable is to alter the \( r_{it} \) from .25 to .50 in intervals of .05. To achieve a level of \( r_{it} = .50 \), the test-taker needs known questions with known characteristics. The higher the item-test correlation, the fewer the number of items that are needed to adequately cover a domain.

---

55. Tinkelman, supra note 19, at 72. When applying the formula to a particular situation in an attempt to calculate the optimal number of questions for a test, there are usually many unknowns. In order to effectively use the formula, you must make assumptions. Telephone Interview with Dr. Julian J. Szucko, supra note 22.

56. See Wood, supra note 15, at 245-47. \( r_{it} \) is defined as “the average item-test correlation in the test (or test part).” Tinkelman, supra note 19, at 71.

57. Telephone Interview with Dr. Julian J. Szucko, supra note 22.

58. However, .05 is unlikely because item writers tend to be inexperienced in their initial writing of the exam. This suggests that the exam should ideally be written so that there is no confusion to the test-takers, for this to occur the items must be well written.
### Table I

<table>
<thead>
<tr>
<th>Formula</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_u = .80$</td>
<td>Reliability of Test</td>
</tr>
<tr>
<td>$r_{it} = .30$</td>
<td>Average Item-Test Correlation</td>
</tr>
<tr>
<td>$n =$</td>
<td>Minimum Requisite # of items/questions per domain</td>
</tr>
<tr>
<td>$r_{it} = \frac{nr_{ii}^2}{1+(n-1)r_{ii}^2}$</td>
<td></td>
</tr>
<tr>
<td>$.8 = \frac{.09n}{1+(.09n-.09)}$</td>
<td></td>
</tr>
<tr>
<td>$.8 = \frac{.09n}{.91+.09n}$</td>
<td></td>
</tr>
<tr>
<td>$(.91+.09n).8 = .09$</td>
<td></td>
</tr>
<tr>
<td>$.728+.072n = .09n$</td>
<td></td>
</tr>
<tr>
<td>$.728 = .018n$</td>
<td></td>
</tr>
<tr>
<td>$40.4 = n$</td>
<td></td>
</tr>
</tbody>
</table>

Note: if $r_{it} = .3$ then approximately 40 items/questions are needed per domain.

This equation is based on an assumption of a single domain. The domain(s) that the professor include(s) in the examination is an individual judgment call. The domain can either be a single broad area of knowledge or multiple areas.  

### IV. THE WOOD STUDY

In 1924, Professor Benjamin D. Wood of Columbia University published the results of a complex study (Wood Study) that he had designed to measure the reliability of both the traditional essay law school exam and a new type of exam, the objective examination. Applying psychometric

---

59. Telephone Interview with Dr. Julian J. Szucko, *supra* note 22.
61. See id. 247-48.
principles, Professor Wood measured the internal consistency of nine items from the same test and found that the traditional law school essay exam had an internal reliability (r) of $r = 0.60$, a level that was statistically unacceptable. For his study, Professor Wood designed tests for three law school courses. Each test consisted of three to four "old type" case method essay questions and either 75, 97 or 200 "[n]ew type" objective questions. These courses were identified simply as I-A, I-E, and II-A. I-A received seventy-five true/false questions, I-E received two hundred true/false questions, and II-A received ninety-seven true/false questions. Professor Wood’s measurement of the internal reliability of these tests revealed that the objective sections correlated with the essay questions given in other tests, thus indicating that the true/false examination measured exactly the same material as the traditional essay exams — and did so as reliably as the essay questions correlated with themselves. According to the Wood Study, Test I-E (two hundred true/false questions) had the highest reliability coefficient: $r = 0.57$. Test I-E also had the greatest correlation with first year law school grades although yielding a slightly lower reliability, $r = 0.52$. In addition, the objective exams correlated more directly with intelligence tests than did the traditional essay exam consisting of three to four essay questions to be done in an hour and a half.

The objective exams also had a greater correlation with course content — yielding an average reliability of $r = 0.66$, compared to the traditional essay exams that yielded an average reliability of $r = 0.54$. Specifically, of the three tests, test I-E, consisting of two hundred true/false questions had the greatest correlation with course content of $r = 0.80$; II-A, consisting of ninety-seven true/false questions, had a correlation of $r = 0.62$, and

---

62. See id.
63. Id. at 247.
64. See id.
65. See id.
67. The average reliability was: $r = .47$. See Wood supra note 15, at 250.
68. The correlation for traditional essay exams averaged: $r = 0.336$. See Wood, supra note 15, at 247, 250.
69. One commentator notes:

For a 1-hour achievement test, the extended-answer essay format cannot provide a very good sample of a content domain, unless that domain is very limited. The multiple-choice format permits from 50 to 60 items to be administered in that comparable 1-hour test. Therefore, the sampling of content is generally greater than with the use of the essay format.

HALADYNA, supra note 29, at 28.
70. See Wood, supra note 15, at 251.
71. See id. at 250. With respect to the use of multiple-choice or true-false questions on law school examinations, Professor Wood indicated that:
I-A, consisting of seventy-five true/false questions, had the lowest correlation of $r = 0.56$.

| Table II
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comparative Reliabilities of Short Essay Exam vs. True/false Exam</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Indicated Courses</strong></td>
<td><strong>Traditional Exam</strong></td>
<td><strong>True/false Exam</strong></td>
</tr>
<tr>
<td>I-E (200 questions)</td>
<td>$r = .53$</td>
<td>$r = .80$</td>
</tr>
<tr>
<td>I-A (75 questions)</td>
<td>$r = .55$</td>
<td>$r = .56$</td>
</tr>
<tr>
<td>II-A (97 questions)</td>
<td>$r = .54$</td>
<td>$r = .62$</td>
</tr>
<tr>
<td><strong>Average Reliability</strong></td>
<td>$r = .54$</td>
<td>$r = .66$</td>
</tr>
</tbody>
</table>

The results of the Wood Study were as follows: (1) objective examinations correlate with the traditional essay examinations as highly as traditional essay exams correlate with other exams; (2) objective examinations are more reliable; (3) objective examinations produce a greater measure of thinking ability; and, (4) two hundred is the optimal number of items/questions needed to insure both reliability and content validity.73

By comparison, an “objectified” essay examination would have to consist of ten long essay questions testing twenty samples of knowledge allowing the student to earn fractional credit toward the total point value of the question (for example, five sub-groups of questions for each of four domains). Alternatively, another comparable “objectified” essay examination would have to consist of twenty long essay questions that each

---

New Type measures what the old type measures and does it with greater reliability. . . . it was also desirable to find if possible just how small a number of questions of the new type would suffice to give a satisfactory degree of significance and reliability. It seems that [two hundred] is the minimum from which we can expect satisfactory results.

*Id.* at 249.


73. *See* Wood, supra note 15, at 254. Again, multiple-choice or true/false questions yielded a reliability coefficient of $r = .80$ *See id.* at 251. Thomas M. Haladnya says that under the correct conditions “[a] 50-item test can produce highly reliable scores.” HALADNYA, supra note 29, at 27.
contain the ten opportunities for the student to earn credit (for example, five sub-groups of questions for each of only two domains). 74

Table III

<table>
<thead>
<tr>
<th>Questions with 6 sub-parts</th>
<th>Exam with 3 Objectified Essay Questions</th>
<th>Exam with 5 Objectified Essay Questions</th>
<th>Exam with 7 Objectified Essay Questions</th>
<th>Exam with 10 Objectified Essay Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 graded sub-parts that grade will be based upon</td>
<td>30 graded sub-parts that grade will be based upon</td>
<td>42 samples that grade will be based upon</td>
<td>60 samples that grade will be based upon</td>
<td></td>
</tr>
<tr>
<td>Questions with 9 sub-parts</td>
<td>27 samples that grade will be based upon</td>
<td>45 samples that grade will be based upon</td>
<td>72 samples that grade will be based upon</td>
<td>90 samples that grade will be based upon</td>
</tr>
</tbody>
</table>

Of course, only the most masochistic law professor would administer a final examination that requires her to compose and to grade ten or twenty long essay examination questions. On the contrary, the typical law school essay examination consists of only two to four standard length questions and maybe four to twelve if the questions are very short. Either way, law professors who grade essay exams as though they were multiple-choice or true/false exams in an attempt to increase their objectivity will rarely, if ever, actually base the grade on a sufficiently large number of items/questions. And although there is merit in the sentiment against subjectivity, Table III shows that merely “objectifying” the typical essay exam does very little to increase its validity. Of course, essay exams that are graded without an attempt to “objectify” them are probably the least reliable and the least valid method of testing law school subjects.

Time is also an important factor in the traditional law school examination. Although some law professors administer take home exams or otherwise allow their students to take exams without time constraints, most law school examinations are strictly timed. Typically, the number of hours allowed to complete the examination will correspond to the number of hours of course credit that the student will earn upon achieving a passing grade on the exam. 75 Here too, there is some evidence that the objective

74. See infra Table III and accompanying text.
75. There are, of course, many variations on this theme. Some students may be allowed more time than others depending upon, for example, physical disability or language barriers. Specifically, at The John Marshall Law School, the director of the LLM for Foreign Lawyers has asked the faculty to allow foreign students addi-
examination format is more reliable than the essay format, "particularly [(as is the case for most law school examinations)] if the administration time is one hour or more."  

V. COMMON OBJECTIONS TO OBJECTIVE EXAMINATIONS

A. "I Do Not Know How to Draft Objective Exams."

It is often said that essay examinations are easier to write, but more difficult to grade. Conversely, objective examinations are thought to be more difficult to write but easier to grade. Consequently, even those law professors who have no other objections to using objective examinations often hesitate to use multiple choice questions because they do not know how to draft them. Law professors receive little, if any, training or guidance for teaching, drafting, and grading exams in other than the "traditional" ways. However, the skills necessary for writing multiple-choice or true/false exams can be learned.

I was fortunate enough to have access to an expert in the fields of training and test preparation who uses the following approach to formulating objective questions. This formula is based on the assumption that the goal of the examination is to test specific material or information that had been assigned to the test-taker and that the test taker had actual access to the material prior to the exam (e.g., textbooks, cases, or lectures). Because the questions must be related to the study materials, they can be easily drafted by relating them back to the material. It is not necessary to limit questions to material actually covered in class. As long as the material was assigned, then the test-taker must be prepared to be questioned about it. To draft the alternatives (e.g., a., b., c.), simply extract any declarative sentence or paragraph from the assigned material that contains any information you intend to test. Next, extract the subject or object directly from the sentence or paragraph. Take the "truest" statement in the subject or object to complete their exams.

76. HALADYNA, supra note 29, at 27. "Generally, . . . about one multiple-choice item [can be administered] per minute. . . . With an essay test, writing for 1 hour or more can be exhausting, and the quality of measurable information may be less dependable than desired." Id. To reiterate:

For a 1 hour achievement test, the extended-answer essay format cannot provide a very good sample of a content domain, unless that domain is very limited. . . . The multiple-choice format permits from 50 to 60 items to be administered in that comparable 1-hour test. Therefore, the sampling of content is generally greater than with the use of the essay format.

Id. at 28.

77. Telephone Interview with Bennie L. Crane, supra note 22.

78. See id.

79. Telephone Interview with Bennie L. Crane, supra note 22.
object of the declarative statement and make it the correct alternative in a multiple-choice question, or the “true” alternative in a true/false question. Notably, each alternative in a multiple-choice question is either “true” or “false” and can be the basis of separate true/false questions.

To develop the other, incorrect answers or distractors, use information that may be relevant to the material but that is not the subject or the object of the declarative statement. Distractors may, however, be drawn from anywhere. As the distractors become less and less relevant to the assigned material, the question and the exam will become less difficult. As the relevance of the distractors to the material increases, the question and the exam will become more difficult.

Of course, many law professors just do not want to spend the time drafting objective tests. One popular solution to this problem is pooling objective questions into exam banks.80

B. “Multiple-Choice Exams Cannot Test My Students’ Writing Abilities.”

Another objection is that objective examinations do not test the students’ writing skills. Even if this was completely true, it would not be a good reason to use an invalid test to evaluate proficiency in the underlying substantive subject upon which the student expects to be tested. Also, there are other more appropriate ways to test writing skills.

The law school curriculum is designed by the faculty and reflects its priorities about which subjects will be either a part of the required course of study or offered as electives to all students who are enrolled in the law school. Therefore, law school faculty have many opportunities to build writing and legal skills training into the law school curriculum. The success of this process depends upon each faculty member’s ability to incorporate these skills into the teaching of their specific subject matter. The testing phase simply follows the teaching phase and the test should measure proficiency in the subject of the class, not the subjects of other classes. This is as true for faculty who teach writing as it is for faculty who teach Property Law or Torts. My Property Law examinations are not expected to test tort principles. Unless the class is specifically designed to do so, I cannot think of a compelling reason why the Property Law exam should test writing skills. This is especially true at schools81 where the writing courses are taught by full-time, tenure-track and tenured faculty. Increasingly, most law schools have professional skills and writing faculty, and formal writing programs that are vested with the responsibility to make sure that the students develop writing and other skills. The continued use of law school essay examinations that are otherwise invalid, despite their

80. See Nickles, supra note 8, at 450-51.
81. Including The John Marshall Law School, where the author has taught for more than a decade.
deficiencies, cannot be justified by the fact that they may test writing abilities. Teaching law students legal writing skills is an important goal to which law schools should dedicate resources. It should not be treated as a collateral matter that is associated with the examination of other subjects, especially when doing so is demonstrably detrimental to the test-taker.

C. "Multiple-Choice Exams Cannot Test My Students' Ability to Spot Issues."

A third criticism of objective law school examinations is that exams in this format cannot test a student's ability to spot issues. In fact many points of law are easier to test using multiple choice test questions. In fact, my initial interest in using alternative formats of examination arose from the fact that essay exams made it too difficult to test my students' ability to spot all of the issues and sub-issues that I wanted to test. One example of this is drawn from my extensive experience teaching at least one section of Property Law every year since 1989. The material on future interests is one of the topics (domains) that I cover in the class. After a few years, it became clear to me that almost all essay examination questions about future interests tested the same issues and elicited the same narrow answer. Without going into boring detail for readers who do not teach this topic, I will simply note the fact that once one adds conditions to a gift of future interest, the answer is going to be that the grantee receives a gift of one of the defeasible fees or executory interests that may or may not be destroyed by the "rule against perpetuities." I found very few exceptions to this fact, even though there were so many other important aspects of the topic that I had covered in class and wanted to test. I was not able to test these other important learning points using essay examination questions, but I was able to do so using objective examination questions. The objective examination format solved this dilemma by making it possible for me to test many more issues than I could using the essay format.

These last two objections have led some law professors, who are concerned about the invalidity of essay exams, to format final exams as part objective and part essay. Again, the most important consideration is the need to include a sufficient number of objective questions; and to grade the essay questions only for writing skills and issue spotting. However, the addition of essay questions will reduce the overall reliability of the exam.82

D. Miscellaneous Objections

Many scholars suggest that essay exams require or test other skills such as "organization" or 'control' of material, . . . 'constructive thought' or

82. Klein & Hart, supra note 11, at 204 & n.8.
"synthesis"" and that it is difficult to truly objectify these functions. Still others argue that objective exams hinder the learning process by causing students to concentrate on getting the points rather than in-depth learning and that students will be more likely to concentrate on memorizing and outlining the material rather than employing thoughtful insight or imaginative interpretations in order to learn the material.

Objective exams test the same thought and organization patterns of essay exams. Objective exams are superior because they can sample proficiency in more domains within a subject in the same amount of time; and are more reliably scored. Studies show that students do not significantly study less or differently for objective exams.

Some professors, especially those who are pre-tenure or tenure-track, feel uncomfortable using objective exams because they fear incurring the wrath of more traditional senior faculty. Ironically the objective examination was the "new type" of exam at the time of Professor Wood's study in 1924, and still is today.

VI. CONCLUSION

It is common knowledge among law school faculty that we are comprised almost exclusively of lawyers who teach, and that we typically have no formal training as educators, nor as testing specialists. Many law professors have only limited experience as practicing attorneys prior to entering academia and a law school.

Upon joining a law school faculty, there is very little training and no training manual. If new law teachers have a reasonably accurate idea of what is expected of them, they will still only rarely have any idea about how to do it. Sometimes law school faculties will provide mentoring to new colleagues. As of the summer of 1999, the Association of American Law School (AALS) Minority Groups has instituted a national Mentoring Program that matches experienced faculty with new law teachers based on teaching and scholarly subject areas of interest. The purpose of this program is to provide a remedy to the problem faced by new law school professors due to the lack of sources of information about teaching law.

83. Kissam, supra note 3, at 442; see also J.A. Coutts, Examinations for Law Degrees, 9 J. SOC'Y PUB. TCHR'S. L. 399, 402-03 (1967).
84. Id.
85. A. Ralph Hakstian, The Effects of Type of Examination Anticipated on Test Preparation and Performance, 64 J. EDUC. RESEARCH 319, 319 (1971).
86. See Wood, supra note 15, at 247-49.
87. See Letter from Professor Robert S. Chang, Loyola Law School-Los Angeles, to Professor Linda Crane, The John Marshall School of Law (July 12, 1999) (on file with author) (thanking the mentors in the AALS' Minority Groups Mentoring Program for participating. The letter includes lists of resources for law teachers of color and a bibliography of 25 law review articles and symposia on the subject of law teaching, pedagogy, and scholarship. The letter also includes the
However, for the most part law professors learn the ropes by trial and error on the job. The job consists of a rather solitary process of selecting casebooks, class preparation, classroom teaching, examination writing, and grading. It is also a common practice for new law teachers to use the same books and methods that their law professors used to teach them when they were law students. Later, we emulate colleagues who are using methods we admire. This means that we teach using the same methods by which we were taught; and we test using the same methods by which we were tested.

Clearly, for something as important as the enterprise of training lay people to become lawyers, this is an unjustifiably, unscientific, and even hap-

following list of topics that the mentors should be prepared to offer assistance and advice to their new law teacher mentees). The letter states:

What mentors can help with and what mentees might ask about (not listed in terms of priority):

(1) teaching:
- discussing possible casebooks and other course materials to use
- discussing teaching methods, approaches to particular course material
- sharing syllabi

(2) scholarship:
- sharing research tips and reading lists
- discussing how to work with a research assistant
- reading drafts and providing constructive criticism of articles and essays
- suggesting other readers, asking other scholars to read a mentee’s work
- sharing information on getting published

(3) networking:
- introducing mentees to other scholars in the mentee’s area of teaching and scholarship (it’s a lot easier to call someone you do not know if you can say, “so-and-so suggested that I call you”)
- suggesting conferences that might be helpful
- suggesting your mentee as a speaker, moderator, discussant, committee member, etc.

(4) career advice:
- discussing how to create a promotion and tenure file
- discussing pitfalls and positive opportunities in faculty politics (a mentor may not know the standards, practices, and political situation at a mentee’s school, but a mentor might be able to suggest questions that the mentee can ask of an inside mentor)
- for those who are mentoring persons who are planning to enter legal academics, discussing the realities of teaching and scholarship, the hiring process, how to “package” oneself for the hiring market, etc.

(5) moral support

Id.

This was what I, and many others I have spoken to about it over the years, have done. It can be very helpful for getting through the first semester or two. Recently, my AALS Minority Groups Mentoring Program mentee at Arizona State University School of Law told me that he is doing the same thing now.
hazard approach. Arguably, it is unconscionably insufficient preparation for fulfilling the part of the job that requires the drafting and the grading of a single examination upon which an entire grade is based.

The traditional model for teaching law school courses is essentially incompatible with the objectives of the law professor who wishes to evaluate student performance fairly and reliably; but who bases that evaluation on the results from the traditional model for testing law students — the essay examination. Under this traditional model for testing, the law school examination is formatted into a small number of comprehensive essay questions that the professor has never taught the student to answer. Essay questions assume that the student's orientation to the material being tested is vastly different than it really is. The traditional model for teaching law students is at odds with the traditional model for testing students' proficiency in the subject — at least if reliability, validity, and fairness are goals.

It is inherently unfair to teach students course material in one way and then to test it in another way. In addition, there is a century's worth of evidence that suggests that the essay question format of the traditional law school examination is highly unreliable due to the large number of subjective factors it allows to influence the final grade.\(^8^9\)

In 1976, the Law School Admission Council published the results of a study by Stephen P. Klein and Frederick M. Hart supporting the idea that factors other than substantive knowledge affect essay grades.\(^9^0\) One factor that correlated highly with success on law school essay examinations was legible handwriting.\(^9^1\) Another leading indicator of higher grades was length.\(^9^2\) Longer answers were viewed by law professors as better.\(^9^3\)

Law schools have an obligation to use the most accurate and internally consistent, or reliable examination methods.\(^9^4\) The essay exam is inherently capricious not only because of the number of subjective factors used in scoring that influence the student's overall grade; but also because they compare law students based on too few samples of each student's knowledge of a given domain of material to be reliable or statistically valid.\(^9^5\)

The traditional law school essay exam is mathematically unsound and

---

90. See Klein & Hart, supra note 11, at 197-98.
91. See id. at 199, 201-03; Church, supra note 11, at 828.
92. Klein & Hart, supra note 11, at 199.
93. See id. at 201, 204-05.
94. See Nickles, supra note 8, at 443-44.
95. See HALDYNA, supra note 29, at 26-18, 34; Wood, supra note 15, at 224-26; Telephone Interview with Dr. Julian J. Szucko, supra note 22.
unable to consistently measure the law student's proficiencies within the
law school's curriculum. This is due to an inability to either accurately
sample the same amount of material or to render the same number of sam-
ples of a given domain of material as an objective exam can within a com-
parable time period. Therefore, single-shot essay exams used to measure
numerous domains of information within each larger law school subject
are notoriously subjective and unreliable. Accordingly, they are also inva-
lid for their intended purpose. This is especially true given the enormous
importance placed on the results of law school essay examinations and
because those results are used to compare students' performances.

The essay exam format is inherently incapable of affording law students
an adequate opportunity to demonstrate proficiency in an entire subject.
It is infeasible for the professor to draft an essay exam that is capable of
sampling a sufficient quantity of information from various the domains of
a complex subject. If the professor were to draft successfully an essay
examination that was lengthy enough to contain enough questions for the
examination to be considered valid, it would be impossible for the student
to actually complete the examination within normal time constraints; and
various physical and psychological phenomenon would hinder the students
ability to perform well during the course of completing such an arduous
task. Critics of essay examinations doubt that their unreliability can be
lowered to a level that makes them valid.

On the other hand, objective exams test the same thought and organiza-
tion patterns of essay exams. Objective exams are "superior because
they can sample a wider area of subject matter in an equal amount of time
and are more reliably scored." Studies show that students do not study
significantly less or differently for objective exams. Many law profes-
sors just do not want to spend the time drafting objective exams. None-
theless, there are solutions to this problem such as pooling objective ques-
tions, teaching the way we test, or testing the way we teach.

96. See HALADYNA, supra note 29, at 26-28.
97. See Wood, supra note 15, at 225-26. Telephone Interview with Dr. Julian J.
Szucko, supra note 22.
98. See supra Table III and accompanying text.
100. Nickles, supra note 8, at 447 & n.120 (citing Bracht & Hopkins, The Com-
munality of Essay and Objective Tests of Academic Achievement, 30 EDUC &
PSYCH. MAN. 359 (1970)). Many other scholars have discovered that "research
evidence does not support the common assumption that essay and objective tests
measure different variables as well." See id. (listing these other scholarly articles).
101. See id. at 447.
102. See id. at 449; see also Hakstian, supra note 85, at 323.
103. See Nickles, supra note 8, at 450-51.
104. Another alternative is a combination of short essay questions and objective
questions.
“Objectifying” the essay exam by grading it as though it were a multiple-choice or true/false examination is better than a completely subjective, ad hoc grading procedure. It does not, however, solve the plague of unreliability and invalidity in law school examinations. Applied mathematical proofs have shown that these problems can be avoided most effectively through the use of the objective, not merely the “objectified” examination format.