

1993

"Alternative" Grading in Large Section Law School Classes

Paul T. Wangerin

Follow this and additional works at: <https://scholarship.law.ufl.edu/jlpp>

Recommended Citation

Wangerin, Paul T. (1993) ""Alternative" Grading in Large Section Law School Classes," *University of Florida Journal of Law & Public Policy*. Vol. 6: Iss. 1, Article 3.

Available at: <https://scholarship.law.ufl.edu/jlpp/vol6/iss1/3>

This Article is brought to you for free and open access by UF Law Scholarship Repository. It has been accepted for inclusion in University of Florida Journal of Law & Public Policy by an authorized editor of UF Law Scholarship Repository. For more information, please contact kaleita@law.ufl.edu.

“ALTERNATIVE” GRADING IN LARGE SECTION LAW SCHOOL CLASSES

*Paul T. Wangerin**

I.	INTRODUCTION	53
II.	THE “BASE GRADE/UPGRADE” SYSTEM	56
III.	MIDTERM EXAMS AND “OBJECTIVE” QUESTIONS	60
IV.	A RADICAL ALTERNATIVE	62
V.	GRADUATING SENIORS	68
VI.	CONCLUSION	72

I. INTRODUCTION

An overwhelming amount of anecdotal evidence indicates that the vast majority of American law school teachers use a grading system in their large section, substantive law courses — i.e. their non-writing, non-clinical courses — that has two principal components. First, that system has teachers using a single, end-of-the-term final exam, an exam that generates the entire grade for the course. Second, that system has teachers primarily, if not exclusively, using subjectively graded essay examinations. Although a number of commentators have suggested the use of alternative grading systems in law school classes,¹ nothing systematic has heretofore been published on this extraordinarily important topic. The present analysis attempts to fill that gap in the literature. It does so by describing several “alternative” law school grading systems;² systems that have all actually been used, or are currently

* Associate Professor of Law, John Marshall Law School; J.D. John Marshall Law School (1978); A.B. University of Missouri, Columbia (1969); Clerk, Honorable Howard C. Ryan, Illinois Supreme Court (1978-1979); Associate Attorney, Winston & Strawn, Chicago (1979-1982). An earlier version of this paper was delivered at the 1992 AALS Conference on Property Law (Spokane, WA).

1. See, e.g., MICHAEL S. JOSEPHSON, *LEARNING AND EVALUATION IN LAW SCHOOL* (1984); Philip C. KISSAM, *Essay: Law School Examinations*, 42 VAND. L. REV. 433 (1989).

2. For a number of reasons, none of the grading systems described herein have been the subject of formal empirical research. Thus, no formal data exists *proving* that any of these systems produces more reliable grades than the traditional law school grading system or better learning than the traditional system.

being used, in large section substantive classes at a single American law school.³

Before the present analysis can proceed, however, two propositions about which informed law school teachers cannot reasonably disagree must be put forward. First, grading is far and away the most distasteful aspect of law school teachers' jobs. Thus, most law school teachers will not give the slightest consideration to the use of teaching/grading techniques that call for a larger commitment of grading time than that already spent. Second, the grading system used in most law school classes, the system that primarily relies on the use of a single end-of-term essay exam, is not consistent with generally accepted theory regarding grading in higher education. This theory indicates that a much better grading system involves frequent testing and frequent feedback.⁴

Despite that fact, however, three things suggest that the following analysis has at least some value.

First, a massive amount of research evidence exists *outside of the law schools* indicating that analogs to the alternative grading systems described herein produce better grading and learning than analogs to the traditional law school grading system. This analogous data, in turn, should be especially valuable in the present context. Lawyers, after all, take great pride in the notion of reasoning by analogy.

Second, long established practice in American legal education indicates that "anecdotal" reports about teaching and learning should be allowed to carry great weight. Countless legal educators, for example, claim that they teach students to "think like lawyers." Virtually none of these educators, however, offer formal proof supporting their claims, nor has anyone asked for such proof. The *Journal of Legal Education*, the principle forum for essays on American legal education, regularly publishes articles that provide no formal proof whatsoever regarding the teaching and learning claims made by the authors of those articles. But again, no one complains.

Finally, the following analysis has some value in connection with the planning of future formal educational research. The alternative grading systems described herein are the results of an enormous amount of informal trial and error experimentation involving a seemingly endless number of "pilot studies." Such studies often produce valuable information and ideas. Further, since the following analysis provides many specific details regarding the alternative grading systems described herein, the analysis itself can be used as a blueprint by researchers and other teachers. Admittedly, blueprints may not be as valuable as the buildings themselves, but construction cannot be done without them.

3. A brief word must be said here about the John Marshall Law School in Chicago at which these alternative grading systems have been used. John Marshall is one of the few remaining "free-standing" (that is, not formally affiliated with a university) law schools in the United States. The academic program at John Marshall, like the academic programs at most other free-standing law schools, is now very similar to the academic programs at most American law schools; however its history, like the history of most other free-standing law schools, is that of a "flunk out/bar exam preparation" school.

The students attending John Marshall are comparable to the students who attend virtually any other "regional" law school in America. Their admission indicators are good, but not great. (The admissions indicators of scholarship students at the school, however, are often extraordinary.) Most of the students who attend John Marshall attended non-prestigious undergraduate institutions—often non-elite state universities—and many are from non-professional families. Finally, most of the students at the school intend to practice law in small firms, in minor governmental entities, or as sole practitioners.

4. The literature on teaching and testing, which universally supports this point, is immense. For easy to read, yet comprehensive discussions, see JAMES S. CANGELOSI, *DESIGNING TESTS FOR EVALUATING STUDENT ACHIEVEMENT* (1990); LOU CAREY, *MEASURING AND EVALUATING SCHOOL LEARNING* (1988); T. DARY ERWIN, *ASSESSING STUDENT LEARNING AND DEVELOPMENT: A GUIDE TO THE PRINCIPLES, GOALS, AND METHODS OF DETERMINING COLLEGE OUTCOMES* (1991); NORMAN E. GRONLUND & ROBERT L. LINN, *MEASUREMENT AND EVALUATION IN TEACHING* (6th ed. 1990); JOHN HEYWOOD, *ASSESS-*

MENT IN HIGHER EDUCATION (2d ed. Wiley 1989); KENNETH D. HOPKINS ET AL., EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT AND EVALUATION (7th ed. 1990); STEVEN J. OSTERLIND, CONSTRUCTING TEST ITEMS (1989). In this writer's opinion, Cangelosi's and Carey's books of the foregoing are the *least* useful to law school teachers. For a more complex treatment of this subject, see ROBERT L. EBEL, ESSENTIALS OF EDUCATIONAL MEASUREMENT (5th ed. 1991); NORMAN FREDERIKSEN ET AL., DIAGNOSTIC MONITORING OF SKILL AND KNOWLEDGE ACQUISITION (1990).

Short discussions of testing theory are contained in ASSESSING STUDENTS, APPRAISING TEACHING (1981); KENNETH E. EBLE, THE CRAFT OF TEACHING: A GUIDE TO MASTERING THE PROFESSOR'S ART (2d ed. 1988); BETTE L. ERICKSON & DIANE W. STROMMER, TEACHING COLLEGE FRESHMEN (1991); Bill J. Frye, *Planning Student Evaluation, Constructing Tests and Grading*, in TEACHING IN COLLEGE: A RESOURCE GUIDE FOR COLLEGE TEACHERS 169-201 (Donald Greive ed., 1989); Barbara S. Fuhrmann & Anthony F. Grasha, *Basic Principles for Evaluating Students*, in A PRACTICAL HANDBOOK FOR COLLEGE TEACHERS 165-92 (1983); PATRICK W. MILLER & HARLEY E. ERICKSON, TEACHER WRITTEN STUDENT TESTS: A GUIDE FOR PLANNING, CREATING, ADMINISTERING AND ASSESSING (1985); Thomas R. Guskey, *Learning and Evaluation*, in IMPROVING STUDENT LEARNING IN COLLEGE CLASSROOMS 31-45 (1988); RICHARD D. KELLOUGH, A RESOURCE GUIDE FOR EFFECTIVE TEACHING IN POSTSECONDARY EDUCATION: PLANNING FOR COMPETENCE 266-83 (1990); WILBERT J. MCKEACHIE, TEACHING TIPS: A GUIDE FOR THE BEGINNING COLLEGE TEACHER (8th ed. 1986); Joseph Lowman, *Evaluating Student Performance: Testing and Grading*, in MASTERING THE TECHNIQUES OF TEACHING 184-209 (1984); David Newble & Robert Cannon, *Assessing the Students*, in A HANDBOOK FOR TEACHERS IN UNIVERSITIES AND COLLEGES: A GUIDE TO IMPROVING TEACHING METHODS 93-125 (1989). For general summaries of theories about teaching, testing, and learning in higher education institutions, see ROBERT J. MENGES & B. CLAUDE MATHIS, KEY RESOURCES ON TEACHING, LEARNING, CURRICULUM AND FACULTY DEVELOPMENT (1988).

The most commonly perceived problems with the traditional law school grading system can be quickly described. The use of a single final exam system prevents teachers from determining, during the course, whether students taking the course are actually learning what the teacher thinks they are learning. The single final exam system does not help teachers determine whether they ought to make teaching adjustments. For a recent discussion of the reasons for, and the nature of, classroom assessment, see THOMAS A. ANGELO, CLASSROOM ASSESSMENT TECHNIQUES: A HANDBOOK FOR COLLEGE TEACHERS (2d ed. 1993); Thomas A. Angelo, *Introduction and Overview: From Classroom Assessment to Classroom Research*, in CLASSROOM RESEARCH: EARLY LESSONS FROM SUCCESS: NEW DIRECTIONS FOR TEACHING AND LEARNING (No. 46) (Thomas A. Angelo ed., 1991). Second, teachers who use a single final exam grading system cannot provide their students with repeated testing and feedback. It is generally acknowledged by most educators that students learn best when they receive repeated tests and repeated feedback. *But see* Steven Hartwell & Sherry L. Hartwell, *Teaching Law: Some Things Socrates Did Not Try*, 40 J. LEGAL EDUC. 509-23 (1990). Finally, a great deal of empirical evidence, outside of the law schools, reveals that the use of grading systems that rely on only a few, or on a single, exam almost inevitably creates "inaccuracy" in grading. Inaccuracy in grading occurs when the final grades generated by a grading system do not accurately reflect students' abilities. Very good law students, for example, might experience serious personal problems just before a final exam, or might have very bad migraine headaches on the day of that exam, and perform poorly. The resulting final grades would not be accurate. Likewise, poor students might have studied only that material which appears on the single final exam and might score well. These grades also would be inaccurate.

The subjective grading aspect of commonly used law school grading systems also produces several problems. First, subjective test grading takes a far greater amount of time than does objective test grading. Since law school teachers often have large numbers of students and rarely have graduate assistants available to help them, the subjective grading of essay tests places an enormous time burden on such teachers. Second, a vast amount of empirical research done outside of the law schools indicates that subjective grading inevitably leads to "unreliability" problems in grading. Unreliability occurs when varying results are obtained by different people grading the same test or when one person reaches different results when grading the same test on several different occasions. For example, if one teacher grades the same subjective exam five different times, that teacher is likely to give several different grades. Likewise, if five different teachers grade the same subjective exam, those different teachers are likely to give different grades even if they totally agree on all substantive points.

It should now be immediately clear that an irreconcilable conflict seems to exist between the two propositions just stated. Most law school teachers will not do anything that increases the amount of time that they spend grading. This is so despite the fact that better learning will almost certainly occur in law school if teachers provide students with something more in terms of testing and feedback than a single final exam. In fact, however, as the present analysis suggests, no irreconcilable conflict exists in this context. It suggests that teachers who use the alternative grading systems described herein, or other comparable alternative grading systems, can provide large numbers of students with frequent tests and feedback and yet devote little or no additional personal time to grading.

II. THE "BASE GRADE/UPGRADE" SYSTEM

Anecdotal evidence suggests that substantial numbers of American law school teachers have in recent years started using a grading system that uses the grade received on a single final exam as the "base grade" for the course, a grade that is then increased (and sometimes decreased) in light of one or more other factors. The most commonly used of those other factors in this context, of course, is class participation. Students who do very fine work in class get "upgrades" from the final exam base grade, usually a half letter grade, or even a whole letter grade.⁵

For many years, Donald Beschle of the John Marshall Law School used a class participation upgrade system such as that just described in his large section property and constitutional law classes.⁶ In recent years, however,

Notwithstanding the problems, the standard law school grading system has defenders who, not surprisingly, defend both aspects of that system. They insist that administering a single final is deeply rooted in history and tradition, which should carry great weight. Further, they argue that a single final exam actually helps identify the best and the worst students. These traditionalists believe that the ability to cope with the enormous pressure, created by a single final exam, and pressure in general, is something that should be factored into law school grades. They also argue that the absence of feedback inherent in the traditional grading system is not problematic. Indeed, some even insist that the absence of feedback is preferred because, ideally, the so-called "socratic method" of instruction widely utilized by American law school teachers places the burden of responsibility for learning on the students themselves rather than on their teachers. Feedback regarding student performance, at least direct feedback, is not a feature of that method. In defending the use of subjectively graded essay exams in law school classes, several arguments are advanced. Most law school teachers do *not* wish to use exams to measure the amount of informational knowledge that students possess. Rather, they wish to use exams to measure students' ability to use very complex skills. These teachers believe objective questions can only measure informational knowledge but not complex skills.

5. Interestingly, a surprisingly large amount of literature addresses the issue of the evaluation of classroom participation. Unfortunately, all of the literature deals with education outside of the law schools. See generally Marilyn Armstrong & David Boud, *Assessing Participation in Discussion: An Exploration of the Issues*, 8 *STUD. HIGHER EDUC.* 33-44 (1983); Edward G. Clarke, *Grading Seminar Performance*, 33 *C. TEACHING* 129-33 (1985); Paul R. Lyons, *Assessing Classroom Participation*, 37 *C. TEACHING* 36-38 (1989); Kenneth B. Melvin, *Rating Class Participation: The Prof/Peer Method*, 15 *TEACHING PSYCHOL.* 137-39 (1988); Gary Westcott, *Teaching and Evaluating Discussion Skills*, 71 *ENG. J.* 76-78 (1982).

6. Beschle scores the class participation of students on an ongoing basis throughout the term. Every

Beschle expanded this upgrade notion considerably. Beschle, like other law school teachers who use the Base Grade/Upgrade grading system, uses the single final exam in his course to create a base grade. Further, like other law school teachers who use this system, Beschle upgrades that base grade in light of class participation. However, he also upgrades the base grade in light of performance displayed by his students on a set of quizzes and on a set of written, take-home mini-exams.

Beschle usually gives about ten to twelve quizzes to his students during the term, devoting no more than about ten minutes of class time to each quiz. Each quiz has one question, a question which Beschle thinks of as the equivalent of "opening" questions for general class discussion. The questions can be relatively easy such as, "What's the key concept discussed in the cases for today?" Or the questions can be relatively hard, "Can you briefly reconcile the holdings of the two cases for this assignment?" Students, who must answer these quizzes in no more than two or three sentences, put their names on their quiz papers and hand them in as Beschle begins regular class discussion.

Beschle then grades these quiz papers at the rate of about 100 papers per hour. He grades them so fast for four distinct reasons. First, Beschle specifically designs questions to produce answers that can be read in just twenty-five to thirty seconds. And he dramatically limits the amount of time that students have to answer the quizzes. Second, Beschle makes no attempt to find fine distinctions when he grades. Rather, he uses a simple zero to three point scale; a scale, incidentally, that is identical to the one that he uses when he evaluates oral classroom participation.⁷ Third, Beschle places no written comments on the papers that he returns other than the numerical score. Fourth, and perhaps most significantly, Beschle grades these quiz papers without endlessly agonizing over the grades he gives. His first impression simply controls.

This last point is crucial. Everyone who has graded papers in law school knows that most teachers spend a large percentage of their overall grading time agonizing over the specific grades that they give.⁸ This kind of agoniz-

time a student speaks in class, Beschle grades that participation on a zero to three scale. He gives: zero points if the answer is not responsive or completely wrong, one point if the answer indicates that the student read the assignment but did not get the basic point, two points if the student answered the question more or less correctly, and three points if the student did a nice job on the question.

An important side benefit accrues because of Beschle's use of this class participation evaluation system. Since students in his classes know that class participation can gain them points, and cannot cost them anything, they tend to participate much more in his classes than in the classes of teachers who do not reward class participation. Further, since students know that better preparation results in additional points, they tend to prepare more.

7. See *supra* note 6.

8. Although few law students seem to realize this, grade conferences between teachers and students regularly force the teachers to lie. This is illustrated by a conference during which a student asks why the

ing slows the grading and makes the whole process a dreaded experience. This is eliminated in Beschle's system. Because of the volume of grades Beschle's system produces, each individual set of grades serves as a check and balance on other sets of grades. Thus, the individual grading mistakes or inaccuracies that Beschle might make because of the speed at which he grades have relatively little overall impact. In the long run, things simply balance out.

Beschle's take home "mini-exams" constitute the second set of grades in his overall grading system. Two or three times during the term, and always at logical breaks in the studied material, Beschle spends about three minutes of class time handing out a take-home mini-exam to his students. This mini-exam consists of a single essay-type question involving a hypothetical factual situation which deals with ideas discussed in the material just completed. Beschle gives students about a week to turn in *typed* answers to these mini-exams, answers that students again identify with their names.

Beschle grades the answers to these mini-exam questions at the rate of about twenty to twenty-five per hour. He grades them so fast for the same reasons that he grades his quizzes so fast, and also because the answers are typed. Beschle designs mini-exam questions to produce answers that can be read in just two to three minutes. Further, Beschle uses a very simple zero to three grading system.⁹ In addition, except in extraordinarily rare situations, Beschle writes nothing on the papers that he returns other than the number score. Finally, and most importantly, Beschle spends no time agonizing over grades that he gives on these mini-exams. He returns the graded papers to

teacher could not have given her just one extra point, and thus a different grade. The teacher then launches into a long discussion as to why that would not be possible. While the defending the grade, the teacher is thinking that he could have given that extra point and that there really is no justification for the lower grade. Consider another regular occurrence: during a conference, a student tells the teacher that a friend answered a particular question in a particular way and got a higher grade (the friend, it seems, has already attended a conference). The teacher then goes into an elaborate description of why this could not be so. The teacher, however, is simultaneously saying to herself that it is possible that the two students answered the question in the same way and that no justification exists for the grade difference.

Many law school teachers try to avoid the problems just described by using elaborate "point" systems. Everything in a question is worth a certain number of points. In effect, these teachers attempt to create objective tests in essay format. Unfortunately, this does not solve the problem, at least not for conscientious teachers. As they total points and assign grades, conscientious teachers realize that in most situations no statistically significant reason exists for giving papers with slightly different numerical scores completely different grades. Thus, they agonize over the borderline grades. These teachers also end up lying. Almost every semester, a student who just missed grading onto law review asks why the cut off for an A was ninety points rather than eighty-nine points. A student recently placed on probation asks why the cut off for a C was seventy points rather than sixty-nine. When confronted with these questions, most teachers give an elaborate explanation which the teacher knows makes absolutely no sense.

9. Because he wants these papers to be the equivalent of three quizzes he uses a nine point scale when he grades them. (His grading probably would go even faster if he used the same zero to three point scale described earlier and then simply multiplied the result times three.) Teachers with access to a computerized spreadsheet could use a zero to three grading scale and then use the spreadsheet to create the appropriate multiplier.

the students about a week after the exam and then spends fifteen to twenty minutes during class discussing the answers.

The first two parts of Beschle's final exam, an exam which is graded anonymously, are extensions of his quizzes and mini-exams. The first part of the final consists of about fifteen or twenty short-answer questions which are extremely similar to the short-answer questions he asks on the quizzes. (Some of these final exam short answer questions, in fact, are duplicates of quiz questions.) The second part of Beschle's final exam consists of two typical hypothetical factual situation questions; questions that are, again, extremely similar to the questions contained on his earlier administered mini-exams. The third part of Beschle's final, a part he uses solely to determine which students will get A's for his course, consists of a single "describe-the-nature-of the-universe" question.¹⁰

As noted earlier, the grade that Beschle gives students on the anonymously graded final exam becomes the "base grade" for the course. Beschle then "upgrades" that base grade one half letter grade for the twenty percent of his students who have the highest grand total scores produced by class participation, quizzes, and mini-exams. Beschle retains at least some degree of anonymity in this process by having the school's records office add the non-anonymous upgrade to the anonymous final exam grade.

Regrettably, despite its strengths, Beschle's alternative grading system also has significant weaknesses. First, the weaknesses: Beschle's grading system involves non-anonymous grading of quizzes and mini-exams, something that clearly violates long-established, and very valuable, law school tradition. Second, Beschle's use of a "grand total" notion in connection with his upgrade system can produce anomalous results. For example, this system can prevent a student who does quite well on mini-exams and quizzes, but who rarely participates in class, from getting an upgrade. Likewise, this system allows a student who is very good in class to get an upgrade despite relatively poor performance on mini-exams and quizzes.

The third weakness in Beschle's upgrade system is perhaps the most significant one. Since the system produces an upgrade for only *twenty percent* of the students in his classes, it provides most of the students with relatively little incentive to work hard on an interim basis as the semester progresses. This is so, of course, because most of the students in Beschle's classes gradually come to realize as the semester passes that they, personally, are unlikely to receive an upgrade. This is especially so for students who do poorly on early quizzes and mini-exams. They stop trying. If they are unlikely to obtain an upgrade, the incentive to study strenuously on an interim basis is lost.

10. Beschle instructs his students not to answer this part of his final if they are not interested in receiving an A for the course.

Despite these shortcomings, Beschle's alternative grading system clearly has great strengths. This system provides both Beschle and his students with a substantial amount of interim feedback regarding the learning that is going on in his classes. Such feedback, in turn, is valuable both for Beschle and his students. Second, Beschle's alternative grading system significantly increases the overall reliability and accuracy of the grades that he gives. Third, and finally, Beschle's alternative grading system accomplishes these things despite the fact that its use requires him to spend little additional time grading.

III. MIDTERM EXAMS AND "OBJECTIVE" QUESTIONS

Many law school teachers insist that any additional time spent grading is too much. These teachers will argue that the benefits to students (and to teachers) produced by additional grading opportunities, no matter how great they may be, simply are not worth the cost. These teachers might be attracted to the alternative grading system used by Allen Kamp, a colleague of Beschle at the John Marshall Law School in Chicago. Kamp believes that his system significantly increases interim feedback for students and teachers while simultaneously *reducing* the overall amount of time that teachers spend grading.

Kamp gives two one-hour midterm exams in his courses; exams that each count as twenty percent of the total grade for his courses. Sometimes Kamp himself grades these exams, but does so with extraordinary rapidity (two hours or less for 100 papers). Other times Kamp assigns the responsibility for initial grading of these papers to staff or student assistants.¹¹ Kamp can do this because his midterms are made up solely of about twenty-five to thirty "objective-type" questions, that is, the true/false, multiple choice, or fill-in-the-blank variety.¹²

11. When Kamp has assistants grade his midterms, he has them circle questionable answers. Kamp himself then reviews all circled answers.

12. Not surprisingly, the literature on objective-type questions, now usually called "selected response" questions, is immense. One of the most interesting, short, recent discussions of this topic is Lewis R. Aiken, *Testing with Multiple Choice Items*, 20(4) J. RES. & DEV. EDUC. 44-58 (Summer 1987). See also Thomas M. Haladyna & Steven M. Downing, *A Taxonomy of Multiple-Choice Item-Writing Rules*, and *Validity of A Taxonomy of Multiple-Choice Item-Writing Rules*, 2 APPLIED MEASUREMENT EDUC. 51-78 (1989). Compare with Randy E. Bennett, *Equivalence of Free-Response and Multiple-Choice Items*, 28 J. EDUC. MEASUREMENT 77-92 (1991); Randy A. Ellsworth, *Multiple-Choice Test Items: What Are Textbook Authors Telling Teachers?*, 83 J. EDUC. RES. 289-93 (1990); Lawrence Feinberg, *Multiple-Choice and Its Critics*, 157 C. BOARD REV. 12-17 (1990); David A. Frisbie & Douglas F. Becker, *An Analysis of Textbook Advice About True-False Tests*, 4 APPLIED MEASUREMENT EDUC. 67-83 (1990); Eric P. Hibbison, *The Ideal Multiple Choice Question*, 22(2) FORUM FOR READING 36-41 (1991) (recent treatments of objective testing). But see OHMER MILTON, *WILL THAT BE ON THE FINAL?* (1982) and OHMER MILTON ET AL., *MAKING SENSE OF COLLEGE GRADES: WHY THE GRADING SYSTEM DOES NOT WORK AND WHAT CAN BE DONE ABOUT IT* (1986) (providing arguments against objective testing).

Interestingly, a number of computer programs exist that can be used to develop multiple choice questions. These programs might be of interest to teachers who wish to develop, and then repeatedly use,

It must be noted that Kamp does *not* believe that objective-type questions can be used to evaluate the complex thinking skills that law school teachers seek to teach in class. Such questions, Kamp believes, can only be used to test for knowledge of information and for very basic skills. Thus, Kamp shares with many law school teachers a general disdain for objective-type questions. Kamp uses such questions on his midterms, however, for two reasons. First, he believes that knowledge of information is extremely important for law students.¹³ Second, he uses these questions because he believes that a *one-way* correlation exists between knowledge of information (and the mastery of simple skills) and the mastery of complex skills. What that means, according to Kamp, is this: students who possess large amounts of knowledge about information and students who display mastery of simple skills may *not* have mastered complex skills. Thus, good grades on tests of knowledge and simple skills will not necessarily be matched by good grades on tests of complex skills. However, mastery of complex skills requires

a lot of multiple choice questions. See Victoria Sturtevant & Bryce Johnson, *Micro-Computer Test-Generation Systems: A Software Review*, 16 TEACHING SOC. 49-54 (1988); Edward L. Vockell & Jane Hall, *Computerized Test Construction*, 62(3) CLEARINGHOUSE 113-20 (1988).

13. Anecdotal experience suggests that a startling discontinuity exists regarding the topic of "informational knowledge" between the views of many law students and the views of many law school teachers. Many law students think that they achieve success principally by memorizing informational knowledge, i.e., rules of law. At the same time, however, many law school teachers insist that informational knowledge is not important at all. It is only important as a tool for learning to think. Regarding law school exams, the students are probably closer to the truth about the role of informational knowledge than many teachers. This is so because skills which the teachers seek to measure cannot be demonstrated in a vacuum. Such skills can only be demonstrated by the manipulation of informational knowledge. A student cannot make a policy argument without knowing facts. Rather, that student can make the argument only if she knows the pertinent rule and the pertinent policy behind that rule. Admittedly, students who possess informational knowledge but few skills will get, at best, mediocre grades on most law school exams. But students who possess skills but little informational knowledge will get, at best, poor grades. *But see* William Perry, *Examsmanship*, 5 HIGHER EDUC. BULL. 133-45 (1977) (Perry recounts the exploits of a Harvard undergraduate who wrote final exam papers without taking the actual courses and got A-'s while many students who took the courses got C's).

The above point has been confirmed by recent work studying the process of thinking. Traditionally, researchers who studied thinking believed, along with John Dewey, that thinking is essentially a generic kind of activity: something that can be done interchangeably in any number of different contexts. Most of these researchers now believe, however, that the quality of thinking may be directly related to the amount of "domain specific knowledge" that individual thinkers possess. They believe good thinkers in particular fields often possess more information than poor thinkers in that same field. See generally *ESSAYS ON THE INTELLECT* (Francis R. Link ed., 1985); *PRACTICAL INTELLIGENCE: THE NATURE AND ORIGINS OF THE COMPETENCE IN THE EVERYDAY WORLD* (Robert J. Sternberg & Richard K. Wagner eds., 1986); *RAYMOND S. NICKERSON ET AL., THE TEACHING OF THINKING* (1985); *THINKING: THE SECOND INTERNATIONAL CONFERENCE* (D.N. Perkins et al. eds., 1987); *TOWARD THE THINKING CURRICULUM: CURRENT COGNITIVE RESEARCH* (Lauren B. Resnich & Leopold E. Klopfer eds., 1989); Robert J. Sternberg & Janet E. Davidson, *A Four-Prong Model for Intellectual-Skills Development*, 22(3) J. RES. & DEV. EDUC. 22-28 (Spring, 1989). For discussions of "critical" thinking, see JOAN B. BARON & ROBERT J. STERNBERG, *TEACHING THINKING SKILLS: THEORY AND PRACTICE* (1987); FRANCIS W. DAUER, *CRITICAL THINKING: AN INTRODUCTION TO REASONING* (1989); BRUCE N. WALLER, *CRITICAL THINKING: CONSIDER THE VERDICT* (1988); JOHN E. MCPECK, *CRITICAL THINKING AND EDUCATION* (1981).

knowledge about information and mastery of simple skills. Thus, good grades on tests of complex skills will always be matched by good grades on tests of knowledge and simple skills. In other words, the mastery of simple things in this context, things that can easily be tested by objective-type questions, are prerequisites to the mastery of complex things.

It should now be clear that Kamp's alternative grading system, if it is based on an accurate analysis of one-way correlation, has extraordinary potential for use in law school classes. Teachers who use this system may well be able to identify very early in the term students who are likely to do poorly on essay-type final exams. (Poor grades on objective-type interim exams, after all, are likely to be followed by poor grades on later essay-type exams.) Yet, teachers who use this system can give as many interim exams as they wish, with just about as many questions as they want, and still spend virtually no additional time grading.

Admittedly, several weaknesses exist in Kamp's alternative grading system, including some fairly serious ones. First, as suggested earlier, if Kamp is correct that only a *one-way* correlation exists between scores on his objective-type midterms and scores on his essay-type finals, the predictive value of the midterms is relatively limited. Students who do well on the objective-type midterms cannot be confident that they will do well on his essay-type finals. Second, since his midterms differ so dramatically from his finals, Kamp cannot really use the midterms as a check and balance of his final exam. Thus, Kamp must continue to agonize over the grades he gives to essay-type answers on his finals. Beschle's quizzes and mini-exams, it should be recalled, serve as excellent checks and balances to his final. Third, since Kamp's midterms differ so greatly from his finals, Kamp cannot use his midterms to give students an opportunity to practice for the final. Again, recall that Beschle's quizzes and take-home exams are direct precursors to his final.

IV. A RADICAL ALTERNATIVE

It should now be clear that neither Beschle's alternative grading system for large section law school classes, nor Kamp's, is ideal. Beschle's system requires a greater commitment of grading time than most law school teachers are willing to make, though far less of a commitment than is initially apparent. On the other hand, Kamp's system requires virtually no additional commitment of grading time. But it does not provide either students or teachers with nearly as much useful information. These facts, in turn, bring this analysis to a description of an alternative grading system that this writer, a colleague of Kamp and Beschle at the John Marshall Law School, uses.

Some of the things that I do in connection with the use of an alternative grading system are similar to things already described. In fact, I pioneered many of the ideas noted above in connection with the grading systems of

Beschle and Kamp. For example, like Beschle, I use interim evaluation tools — quizzes and midterms — that are direct precursors to a final exam. Further, like Beschle and Kamp, I use several different kinds of questions. In addition, like Kamp, though not Beschle, I count the results on midterms and quizzes as formal parts of the overall grade rather than simply as components of a potential upgrade. The scores on each midterm, and the quizzes immediately preceding those midterms, count as twenty-five percent of the total grade for the course. In addition, like Kamp, I grade quizzes and midterms anonymously.¹⁴

Admittedly, several things that I do in connection with the use of my alternative grading system vary somewhat from the things that Kamp and Beschle do in connection with the use of their systems. For example, I record scores for quizzes and midterms, and sometimes even scores for individual questions on quizzes and midterms, on a computerized spreadsheet. (Beschle and Kamp do all of this by hand.) By using a computer, the totalling process is simplified.¹⁵ And, by doing this, simple statistical tests on the results of quizzes and tests are easily conducted. In addition, I systematically use my grading system to encourage students to engage in ongoing review.¹⁶ Beschle and Kamp use their grading systems primarily, if not exclusively, to generate feedback for themselves and their students. Third, I figured out a logistically simple way to give midterms and finals that contain both closed book and open book portions.¹⁷ Beschle and Kamp give in-class exams that

14. Kamp and I do this by having the records office at the school assign anonymous grading numbers to students at the beginning rather than at the end of the term. Students in our classes then use these numbers, instead of their names, on all quizzes and midterms.

15. The use of a computerized spreadsheet can also make it very easy to use different types of questions on the same exam. For example, a teacher gives an exam that consists of twenty five objective questions (worth 1/2 of the grade) and an essay question (worth 1/2 of the grade). The teacher gives one point for correct answers to objective questions and no points for incorrect answers. Further, the teacher scores essays on a scale of 0 through 5. The spreadsheet simplifies the totalling up process. First, the spreadsheet finds the average score achieved by all of the students on the objective questions and the average score for the essay. Then, the spreadsheet divides the average on the objectives by the average on the essay. This factor is then multiplied against each individual student's score for all objective questions. The resulting number is a score for the objective questions that is the equivalent of the score for the essay. Hence, the scores on these two different parts of the exam have been equalized.

16. The grading system is used to force ongoing review in several ways. For example, two of the five questions on the quizzes generally deal with the current assignment while one question deals with an important point made during the previous class. Finally, two of the questions deal with important points made substantially earlier in the course. Further, previous quiz questions are often used on midterm exams and earlier midterm questions are used on later exams.

17. When the midterms and finals begin, both Part I, the closed book objective question part of the exams, and Part II, the open book essay question part, are distributed. Students can begin working on Part II as soon as they complete Part I. At this time, two bluebooks are distributed and students are told to put the answers to Part I in one bluebook and the answers to Part II in the other bluebook. Finally, students are told that they must keep their own books and notes closed for the first hour of the exam, or until an announcement is made that they can open those books and notes. I collect the blue books containing Part I answers before announcing that books may be opened.

are either entirely open book, or entirely closed book. Fourth, unlike Kamp, who uses objective-type questions only to test for knowledge of information and mastery of simple skills, I use objective-type questions to test for quite complex skills.¹⁸ Indeed, major portions of my quizzes and exams are now made up of objective-type questions. Fifth, unlike Kamp and Beschle, who devote relatively small amounts of class time to midterm exams, I devote a great deal of time to such exams. I give a three hour midterm during the fifth week of class and another three hour midterm during the tenth week of class. No regular classes are held during the weeks of the midterms. Finally, I combine a basic quiz/midterm grading system with a somewhat unusual "upgrade" system for class participation.¹⁹

None of the things just described, of course, are particularly "radical." Several things about my system, however, things not yet described, truly are radical. But, thankfully, these things also are strictly optional, things that teachers need not employ in connection with my basic system. Indeed, for a number of reasons, I myself, have temporarily abandoned the use of some of the options described below.²⁰

18. For discussions of methods for developing objective-type questions that test high level thinking skills, see HARRY G. MILLER ET AL., *BEYOND FACTS: OBJECTIVE WAYS TO MEASURE THINKING* (1978); Gale H. Roid & Thomas M. Haladyna, *Items for Prose Learning*, in *A TECHNOLOGY FOR TEST ITEM WRITING* 91-114 (1982); Reed G. Williams & Thomas M. Haladyna, *Logical Operations for Generating Intended Questions (LOGIQ): A Typology for Higher Level Test Items in A TECHNOLOGY FOR TEST-ITEM WRITING* (Gale H. Roid & Thomas M. Haladyna eds., 1982); Gale H. Roid & Thomas M. Haladyna, *The Instructional Quality Inventory*, in *A TECHNOLOGY FOR TEST-ITEM WRITING* 187-200 (1982); see also Mark A. Albanese & Richard M. Jacobs, *Reliability and Validity of a Procedure to Measure Diagnostic Reasoning and Problem-Solving Skills Taught in Predoctoral Orthodontic Education*, 13(4) *EVALUATION AND THE HEALTH PROF.* 412-24 (1990); Stephen P. Norris, *Can We Test Validly for Critical Thinking?*, 18(9) *EDUC. RESEARCHER* 21-26 (Dec. 1989); Steven J. Peitzman, *Comparison of "Fact-Recall" with "Higher-Order" Questions in Multiple-Choice Examinations as Predictors of Clinical Performance of Medical Students*, 65(9) *SUPPL. ACAD. MED.* S59-60 (Sept. 1990) (discussing the use of objective-type questions in high level learning situations).

19. Like Beschle, I keep track of each student's class participation on an ongoing basis, using a zero to five scale rather than Beschle's zero to three scale. At the end of the term, however, each student is given a grade for class participation, a grade consisting of a plus (good), a check (average), or a minus (poor). (About 20% of the students are given pluses, and about the same percentage minuses; the rest get checks.) Further, unlike Beschle, who arrives at his class participation scores quantitatively (i.e., by adding up scores from individual days) I use a qualitative method for determining final class participation scores, (i.e., just getting a general "feel" for how students have done). I send a list of all of the students, with checks, pluses or minuses next to their names, to the school records office. That office converts the list of names to ID numbers used by the students on quizzes and midterms and returns the list. Thereafter, I use this list as a cross-check on the accuracy of my exam/quiz grading. Further, this list is used to give final letter grade upgrades (and downgrades) for class participation to students in borderline letter grade situations.

20. When I first used this now optional part of my alternative grading system, I simultaneously used it in three separate sections of my Contracts II class. As expected, the initial reaction was intensely negative. Unfortunately, allegations of cheating were raised following the first midterm exam that term. Those allegations, in turn, caused a large number of students in one of the three sections to stage a full scale revolt. (Students in another of the three sections also rebelled, but to a much lesser degree. And students in the third section enthusiastically endorsed the use of this option.) For a number of reasons, I abandoned use of

The most radical (and thus the most optional) part of my alternative grading system rests on three generally accepted notions in education theory. First, teachers can design tests, even rather complex tests, that can be *reliably* graded by third party graders. As noted earlier, reliability in this context means that different graders give the same grade to the same paper and that the same graders give the same grade to the same paper at different times. This notion, of course, lies behind Kamp's partial reliance on objective-type questions. Second, feedback that follows quickly after performance on quizzes and exams does much more good than feedback that comes after a considerable delay. Third, and finally, students who evaluate other students' work learn a great deal about their own skills and knowledge while engaging in that evaluation.

The most radical (and thus the most optional) aspect of my alternative grading system should by now be obvious. When I use this option — and, as noted earlier, I have temporarily abandoned its use — I have students themselves serve as the initial graders of each other's work.²¹ This option, of course, reserves *final* grading decisions for the professor. I use this option with quizzes, and either collect students' papers and then redistribute them, or simply ask students to exchange papers. Then I read the answers to the questions and ask students to grade each other's work using a scale of "1" for correct and "0" for incorrect. I tell students that if they are uncertain whether an answer to a particular question is correct or incorrect, they should score it as best they can, then circle the number of the answer. Then I collect the now-graded papers, review the circled answers, enter the scores on a computerized spreadsheet and post a copy of that spreadsheet.

I use a similar process when I use the student grading option in connection with midterms. When using this option, I give a two hour midterm and use the third hour for grading. Immediately after students complete the different parts of the midterms — the first part is closed book and contains about twenty-five to thirty objective-type questions and the second part is open book, and contains a single essay-type question — I collect the bluebooks and redistribute them to the students in my class. The bluebooks, of course, contain students' anonymous numbers and not their names. Grading then proceeds, at least for the objective-type questions, as with the quizzes. I read the answers to those objective questions, and students score each other's

the student grading option during that semester for all three sections. Further, I have been, at least until now, reluctant to resume the use of it.

21. Not surprisingly, there is at least some discussion of peer grading in the education literature. ROLAND HUFF & CHARLES R. KLINE, *THE CONTEMPORARY WRITING CURRICULUM: REHEARSING, COMPOSING, AND VALUING PAPERWORK!* 158-64 (1987); Stuart A. Fry, *Implementation and Evaluation of Peer Marking in Higher Education*, 15(3) *ASSESSMENT & EVALUATION IN HIGHER EDUC.* 177-89 (1990); Madeleine Hunter, *Pare Down Your Paper Work!*, 93(5) *INSTRUCTOR* 74-75 (Jan. 1984); James W. Ney *Collaborative Learning in University Grammar Courses*, 15(2) *INNOVATIVE HIGHER EDUC.* 153-65 (1991).

work with a 1 or a 0. Further, students circle ambiguous answers. Finally, I collect the scored papers, review the circled answers, enter the totals on a spreadsheet, and post a copy of the spreadsheet.

Note carefully two critically important technical points regarding the student grading option just described. First, to keep this grading process from bogging down during class time because of students questions, I categorically refuse to answer questions during the grading period itself about the answers I give (“Wouldn’t it also be correct to say. . .”), or about individual students’ responses to my questions (“This student wrote. . . Is that correct?”). Rather, I tell students who have complaints about the answers, or about the scores that they receive from fellow students, to submit complaint notes to that effect after the papers have been collected. The complaints are required to contain students’ numbers and not their names. Then I review these complaints — and there are surprisingly few of them submitted — and make appropriate adjustments. Second, although I believe that relatively few students cheat when they grade each other’s work, I use a number of simple techniques to reduce or prevent cheating. For example, at the beginning of the term, I ask the students to create an honor code committee, a committee that is charged with investigating allegations of cheating. Further, to prevent students from exchanging quiz papers only with their friends, I regularly change the pattern for exchanging such papers. By doing this, an environment is created wherein students cannot count on their quiz papers being graded only by friends. One day, for example, students may be asked to pass quiz papers to the left. The next day they may be asked to pass quiz papers to the right, and then to the right a second time, or even a third time. Sometimes, quiz papers are simply collected and then redistributed. In addition, I regularly run spot checks on graded quiz and exam papers. If papers are wrongly graded, cheating *might* have occurred. Likewise, scores received by individual students on quizzes and midterms are regularly cross-checked. If major discrepancies exist, cheating *might* have occurred. Finally, and perhaps most significantly, when using this grading option I make it absolutely clear to the students that I will seek to have cheaters expelled from school.

It was noted earlier that the student grading option just described is easily used in connection with objective-type questions. No ambiguity exists regarding the answers to multiple choice, true/false, matching, or fill-in-the-blank questions. Further, relatively little ambiguity exists regarding answers to carefully drawn “short answer” questions. Obviously, the vast majority of complaints that will arise in connection with grading answers to these kinds of questions will involve the specified answer itself. Common sense alone, however, suggests that none of this applies to essay-type questions. Such questions, it seems, cannot possibly be graded by third-party graders, let alone by student graders. Ironically, common sense is wrong in this context.

For many years, experts on educational testing and measurement have

been fascinated by the notion of "holistic" grading systems, systems that can be used in connection with the evaluation of essay answers.²² Holistic grading involves two distinct steps. First, when teachers or testers write test questions themselves in light of these systems, they must simultaneously create a number of different "model answers" or "grids" for the questions, or for the different parts of the questions. When models or grids are used, teachers either create precise examples of good, average and poor work, or they otherwise define different levels of quality. Second, graders then compare the essays being evaluated, or the parts of the essays being evaluated, to the various models or grids and determine which model or grid most closely matches the work being evaluated. Once the graders make such a match, they simply assign the point total that corresponds to the matching model to the work being evaluated.

This is not the place for an extensive discussion of holistic grading techniques. Rather, this is the place simply to note that holistic grading of essay papers accomplishes three very important things. First, it generates remarkably reliable grades (reliability meaning that the same grader gives the same grade at different times to a single paper and that different graders give the same grade to that single paper). It does this, of course, because it turns what is usually a completely subjective process into something that is relatively objective. Second, the use of holistic grading seems to significantly speed up the grading process itself. Holistic graders need not be concerned about every fine point in an essay, nor about totalling up points for each individual matter addressed. Rather, holistic graders need only be concerned about the general form and content of the overall essay being evaluated, or of the overall part of that essay being evaluated. Obviously, this overall approach allows for much faster grading than a step-by-step approach. Third, and finally, holistic grading systems can easily be used to generate reliable grades in connection with the evaluation of high order thinking skills. This can be accomplished because the "models" created for use in connection with such systems can easily be structured to contain differing quality examples of

22. Virtually all of the books on testing cited earlier have extensive discussions of essay-type questions, questions that experts now generally refer to as "constructed response" questions. (Objective-type questions, as noted earlier are now generally called "selected response" questions.) The best works on the *grading* of essays, however, are books from the field of composition instruction. The most widely cited work in this context is EDWARD M. WHITE, *TEACHING AND ASSESSING WRITING* (1985). See also JANICE M. LAUER & J. WILLIAM ASHER, *COMPOSITION RESEARCH: EMPIRICAL DESIGN* (1988). See generally HUNTER M. BRELAND ET AL., *ASSESSING WRITING SKILLS: RESEARCH MONOGRAPH NUMBER 11* (1987); LESTER FAIGLEY ET AL., *ASSESSING WRITERS' KNOWLEDGE AND PROCESSES OF COMPOSING*; RONALD HUFF & CHARLES R. KLINE, *THE CONTEMPORARY WRITING CURRICULUM* (1987); Stephen N. Tchudi, *Evaluating Writing*, in *TEACHING WRITING IN THE CONTENT AREAS: COLLEGE LEVEL 51-62* (1986). But see ERIC D. HIRSCH, *THE PHILOSOPHY OF COMPOSITION 177-78* (1977) (critical of holistic evaluation); Donald A. Daiker, *The Student Essay as Dubloon: Discrepancies in Holistic Evaluation*, 7(2) *J. TEACHING WRITING* 127-41 (1988).

those skills in action.

The essay question student grading option of my alternative grading system can now quickly be described. When I use this option, I simultaneously write exam questions and several different "model" answers. These models display differing levels of knowledge and skill application. Then, immediately after students answer the essay questions on the midterm exams, I collect those answers and redistribute them to students. Scoring models are then distributed and students are asked systematically to score each other's work.

It must be noted forcefully that when I use this particular grading option, I later independently re-score all of the essay answers, for three principal reasons. First, regardless of how reliable holistic scoring might generally be, students themselves simply cannot be counted on, at least early in a the semester, to do sufficiently precise work in scoring their fellow students' essays. Second, even if student scoring *actually is* reliable, and anecdotal evidence that I have obtained suggests that it is, students themselves simply will not believe that it is. Thus student sensibilities, no matter how inaccurate, must be accommodated. Third, given the seemingly infinite capacity of law students to be inventive, law school teachers simply cannot create scoring models that match everything that students are likely to submit. No doubt a few answers simply will not fit any prescribed pattern.

Use of the student scoring option just described is likely to produce two things. First, use of that system is likely to produce extraordinary controversy. (Such controversy, incidentally, is the principal cause of my own temporary abandonment of the whole student grading option.) Second, but not quite so obviously, the use of a student-grading option in this context is likely to produce extraordinarily powerful learning by students. Students who initially grade each other's essay papers gain immediate feedback regarding their own displays of high-level thinking skills. This occurs because students who are grading other students' work actually spend much of the grading time actually thinking about the strengths and weaknesses of the work that they submitted. Further, and perhaps more significantly, the use of a student grading option in this context allows individual students to receive at least some degree of individualized analysis of their own work.

V. GRADUATING SENIORS

Anyone who has taught students who are about to graduate knows that most of these students do not study at all during the course of the term itself and only in token amounts before final exams. Admittedly, a few graduating seniors at the top of their classes continue to struggle even at this late date for higher rankings. And a few students at the bottom of their classes struggle to graduate. For the most part, however, students who are soon to graduate from law school simply cannot be bothered with studying. This fact, in

turn, generates an intriguing problem. Is it possible to get graduating seniors to study in law school, either during the term itself or for final exams?

I use a "yin and yang" approach to solve this problem in my "Remedies" classes. The yin involves an extremely relaxed approach to classroom activity itself. Class participation, for example, is purely voluntary. Further, assignments for individual class sessions are extremely brief. In addition, I videotape all class sessions and make the tapes available to students either for library or home use. Finally, I am quite relaxed when it comes to keeping attendance records. All of these things, of course, appeal tremendously to upper division and graduating seniors.

The yang in the course involves the use of alternative grading procedures. During the first four or five weeks of his class I deliver lectures to the class that summarize the entire field of remedies. I call this series of lectures a "short course" in remedies. As soon as this short course is completed, I give a multiple hour, 100 question, objective type test on all of the material covered. This test, which students take during a week in which the class does not meet for regular sessions, counts for twenty percent of the course grade. Since this test uses nothing but objective type questions, I grade it and return answer sheets to the students within a few days. I then open what I call the "whining window." While the whining window is open, individual students can suggest *in writing* alternative answers to individual questions which they believe should have been accepted. Then, *if two things occur*, I give objecting students additional credit. First, statistical analysis of the question involved must show that questions involved in fact produced problematic responses.²³ Second, the arguments that students submit regarding individual questions must themselves be persuasive.²⁴

Two additional things must quickly be mentioned about the "whining window." First, because I want individual students to carefully review their own work on past exams, I give additional points *only* to students who submit individual objections. Although I may conclude that a particular question was really a bad question, I give additional credit for that question only to students who submitted objections to it. Second, I require students who submit objections to submit two *alternate* objective-type questions for each objection submitted. In effect, therefore, students are "charged" for submit-

23. Although initially copies of answer sheets are scored by hand using a "template," the original answer sheets are sent out to be "scanned." Scanning then produces statistical data about the test itself, and about individual questions on the test. Three kinds of statistical data about individual questions are very important in this context. First, the "difficulty" of a question reveals the percentage of students who got it right and wrong. Second, the "discrimination" of a question describes the ability of the question to differentiate between top students and poor students. Third, "response" data reveals percentages of top scoring students who chose alternative answers to particular questions.

24. Absent this restriction, some students will file cursory objections to every question they got wrong.

ting objections. This is done for two reasons. First, this charge forces the students themselves to think about the test writing process. Such thinking almost certainly will end up improving performance on test-taking. Second, this charge tends to discourage students from filing frivolous objections.

After scores are totaled for this first exam, and after objections to it are logged, scores are posted. Grades, however, are not posted. Rather, scores from highest to lowest are listed. Students then know where they stand in relation to each other — a line is then drawn to indicate the point at which I think performance was unacceptable. (Usually this line is drawn about a third of the way up from the bottom of the list.) Below this line is the “danger zone.” Students whose scores are in the danger zone are potentially at risk of *failing* the course.

After administering this first major exam, I embark on the second part of the course; which is about four weeks long. During this second part, I work through a small number of individual appellate court opinions dealing with substantive issues already covered during the short course. At the end of this discussion of appellate court opinions, I give another 100-question, multiple-hour, objective type exam. This second test, which again is worth twenty percent of the total grade for the course, re-examines students on materials covered during the short course. Further, this second exam tests students on newly studied material.

The grading procedures are the same for this second exam as for the first. The exams are graded and copies of the answer sheets are returned to students. Then the whining window is opened. After objections are logged, scores are listed for this test, plus cumulative scores. Finally, on the list of cumulative scores the new “danger zone” is listed.

The last segment of the class involves the use of problems. Not surprisingly, solutions to these problems can only be developed if students understand the substantive ideas discussed in the short course segment of the course. And, again not surprisingly, solutions to these problems can only be developed if students understand the ideas contained in the appellate court opinions discussed in the second segment of the class.

The course ends with a flourish. During the last week of class, I give what is called a “pre-final” exam. Like all of the exams, this exam is several hours long and has 100 objective-type questions. This exam, which covers materials from all three segments of the course, is worth thirty percent of the total grade. After the scoring and whining window procedures are complete, the scores are posted. More importantly, at this time I post “tentative” letter grades for the overall course. These tentative grades are the grades that would be given if based on the three tests already administered.

Not surprisingly, the “tentative” grades send a strong signal to students who have done poor work in the class. This is done by giving a large number of students very low tentative grades. In Spring 1993, for example, I

gave tentative letter grades of F to almost a third of the ninety students in the class. "You are going to fail this course," I told the students, "unless you make major, major changes in the way you are studying."

It hardly need be said that I well know that I cannot *actually* fail thirty percent of the students in a class, particularly if many of those students are graduating seniors. The Dean and the faculty simply would not allow this to occur, at least not on a regular basis. Further, and more importantly, it hardly need be said that my students know that I cannot fail large numbers. Students, after all, often understand quite well the politics of law school teaching. Thus, the low tentative grades following the pre-final seem to be nothing but a transparent bluff. However, the bluff problem has been solved, but in a round about sort of way. It is done in a manner that allows me to deal with a another serious law school grading problem.

I believe, and share these beliefs with students, that hard-working students sometimes simply cannot do well on some individual teacher's exams. These students have studied hard and understand the material. But, nevertheless, they simply cannot communicate with these particular teachers on exams. No good purpose is served if teachers give these students poor grades. The "safety net" is the answer to this problem. It *guarantees* hard-working students a passing grade for the course. Students must do two things to qualify for the safety net. First, students must submit substantial written projects; projects that can only be completed if students study the material involved in a comprehensive and systematic manner. Second, these students must do "reasonably acceptable" work on the final exam. This second requirement, and its studied ambiguity, eliminates the possibility that students will submit safety net projects and then simply blow off the final.

Obviously, the safety net provides security to hard-working students who, for whatever reason, cannot seem to do well on my exams. Thus, if it did nothing else, it would be a worthwhile addition to the panoply of alternative grading systems. But, the safety net idea also serves another purpose, a purpose related to the bluffing problem already noted. That bluffing problem arises because everybody knows that teachers cannot *actually* fail large numbers of graduating seniors. The safety net eliminates the possibility that hard-working students will get failing grades in my classes. Thus, almost by definition, the only students who can get failing grades in the class are lazy students. And everybody knows that teachers can fail lazy students with impunity. In effect, therefore, the existence of the safety net gives increased credence to the threat provided by low tentative grades.

Certain technical details regarding the safety net must yet be mentioned. First, the students can begin doing safety net projects at any time during the semester. However, the earlier they start, the less work they will have to do and the higher the "guaranteed" grade that they will get. For example, a student who falls into the danger zone after the first exam and then immedi-

ately starts doing safety net projects may be required to do a relatively small project and may get a C+ guaranteed. Conversely, a student who waits until after tentative letter grades are posted will be required to do a very large project and may get nothing better than a guaranteed D or D+.

All of this leads up to my Final Exam. The final provides a chance for redemption for students at the bottom of the class and a chance for improvement for students in the middle. Like all of the other tests, the final is several hours long and is made up of 100 objective-type questions. Further, like all of the tests, I score this test manually, obtain statistical data on it, and then open the whining window. Finally, like the pre-final, the final is potentially worth thirty percent of the total grade.

Overall totals are calculated in a straight-forward manner. First, I "standardize" the scores from the different tests. (This must be done to avoid the "weighting" problems described in the "Grade Conferences from Hell" paper.²⁵) Second, I total the scores on the two interim exams — each worth twenty percent and the scores on the final or pre-final — each worth thirty percent.

One last thing is done before I submit the final grades to the records office. I contact students to whom I anticipate giving grades of D or F by certified letter. Usually these are students who did not take advantage of the safety net, or students who did safety net projects but scored below the "reasonably acceptable" performance level on the final. In these letters, the students are told what kind of additional projects they will have to do to obtain passing grades. I do not do this because I am a "softy" when it comes to failing grades. Rather, I do this because I realize that political constraints make it impossible to actually fail large numbers of students. Since I cannot actually fail many students, and since students know this, I can at least make sure that students who pass the class do a minimum amount of work.

VI. CONCLUSION

Several summary points can now be made about my own two alternative grading systems, and, concomitantly, about the alternative grading systems of Beschle and Kamp. First, my various "safety net" procedures may well produce more studying and learning in classes from poor students than the traditional approach of threats of failure. Students know, after all, that most such threats are merely bluffs. Second, as suggested repeatedly herein, the student grading procedure that I use in some of my classes is a very powerful grading and teaching tool. In fact, it is probably the most powerful such tool described herein. However, as also repeatedly suggested herein, the actual

25. Paul T. Wangerin, *Grade Conferences From Hell: Measurement Error in Law School Grading* (submitted for publication).

use of student grading in large section law school classes may well produce tremendous controversy. Thus, law school teachers should only consider use of the student grading option described herein if they have tremendous confidence in their ability to create confidence in their students and if they have tremendous job security. Third, teachers who use short quizzes and long midterms can provide both themselves and their students with almost continuous interim feedback about student learning. And teachers who use such quizzes and midterms can also force students to engage in ongoing review. Such interim feedback (and review) almost certainly will improve long-term performance. Fourth, teachers who give two or more midterm exams, either in-class or take-home, can provide students with an excellent opportunity to “practice” the kinds of things that they will be required to do on final exams. Such practice almost certainly will improve final exam performance. Fifth, teachers who use any of the alternative grading systems described herein are likely to produce final grades for individual students that are more reliable and accurate than grades produced by the traditional law school grading system.

It was noted at the outset of this analysis that a seeming conflict exists between sound teaching and testing theory — which requires frequent testing and feedback — and the incredible dislike displayed by most law school teachers regarding the grading of student work. An additional point made at the outset indicated that alternative grading systems described herein, and other comparable ones, allow teachers who use them to provide students with substantial amounts of interim testing and feedback. Yet these grading systems require little or no additional commitment of grading time from teachers. Regrettably, no proof was offered herein on this second point; no proof, that is, establishing that these alternative grading systems can in fact be used with little or no additional expenditure of *overall* grading time. That absence of proof, in turn, brings this analysis to its final point.

Interestingly, proof one way or the other on the point just made can be individually developed by individual law school teachers, at least by teachers who wish to conduct a simple “action research” experiment. Action research in educational settings, which is conducted informally by classroom teachers rather than formally by educational researchers, involves methodologically simple experiments and informal or seat-of-the-pants data analysis.²⁶ Such an experiment on this issue can easily be described. First, a teacher wishing to gather informal proof on this issue could total the number of hours spent grading final exams at the end of a recent semester (the “control” semester). Second, during about the fifth week of the “experimental” semester, this

26. The topics of “action research” generally, and informal classroom experiments specifically, are discussed at length in PAUL T. WANGERIN, *Action Research in Legal Education* (submitted for publication).

teacher could create and administer a midterm exam that could be graded in one-fourth of the total amount of time already determined. This exam might contain some of the speed-grading techniques described herein. Third, during the tenth week of class, this teacher could then create and administer a second midterm exam, again an exam that could be graded in about one-fourth of the original time. Fourth, during the experimental semester, this teacher could create and administer a final exam that could be graded in one-half of the original time. That final exam would contain at least one question similar in form to the questions on the exam from the control semester and plus some questions similar in form to the ones contained on the midterms.

When finished; the teacher would *informally* analyze the data produced on three different issues. Such informality is appropriate, after all, because this is an "action" research experiment and not a formal educational study. A teacher conducting this little experiment could first informally decide whether the students learned better during the experimental semester (the one just completed), than during control semesters, (previous semesters). The teacher can make such an informal judgment simply by comparing the performance on the similar final exam questions from the two semesters. If better learning seems to have occurred during the experimental semester than during the control semester, and if no outside factors seemed to have played a role in that improved performance, then the teacher can conclude, at least informally, that the alternative grading system produced better learning without an increase in grading time. In other words, a benefit was produced without a cost. Second, a teacher conducting this little experiment could informally decide whether there is "more comfort" with the grades given during the experimental semester than with the grades given during the control semester. In other words, the teacher could informally decide whether there was cause for belief that the grades given during the experimental semester produced a more accurate picture of students' abilities than the grades given during the control semester. If this occurred, and, if student learning during the experimental semester was no worse than student learning during the control semester, then use of the alternative grading system seems to have produced a gain with no cost. Again, since this is an action research experiment, internal and external factors that could also have caused that result are discounted. Finally, the teacher could determine, perhaps by conducting a simple poll of their students, or by adding a question or two to the student evaluation forms handed out at the end of the term, whether the students themselves liked the grading system used in the experimental semester better than the grading system used during the control semester. If students liked the alternative grading system better, and if no worse learning and grading accuracy occurred during this semester than during the previous one, then, again, use of the alternative grading system produced a gain without a cost.

One final point must now be made. Many law school teachers will im-

mediately insist that the experiment just described has several serious methodological flaws. In fact, these criticisms are appropriate. Action research experiments such as the one just described often contain some methodological problems. And yet, experts agree that such experiments have considerable informal value.²⁷ An interesting issue does arise, however, if methodological complaints are raised about this experiment. Critics who attack the methodology of the experiment implicitly acknowledge that empirically valid judgments can in fact be made about the effectiveness, or lack thereof, of different kinds of teaching and testing techniques. In effect, these critics implicitly acknowledge that the teaching and testing that goes on in *their* classrooms can be measured, at least to a certain degree, by empirical research.

That fact, in turn, creates a delicious irony. As noted earlier, an overwhelming amount of empirical evidence exists outside of the law schools suggesting that the traditional law school grading system is far less educationally sound than the "alternative" grading systems described herein. As a result, reasonable people surely might conclude in this context that the burden of proof rests on the defenders of the traditional law school grading system rather than on the proponents of its alternatives. But, once the burden of proof shifts in this way, the defenders of the traditional system must themselves devise experiments proving that their system is, at least, no worse than the systems proposed herein. Then the battle regarding law school grading really will be joined.

27. *Id.*

