

University of Denver

From the Selected Works of Arthur Best

February 29, 2008

Student Evaluations of Law Teaching Work Well: Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree

Arthur Best, *University of Denver*



Available at: https://works.bepress.com/arthur_best/1/

**STUDENT EVALUATIONS OF LAW TEACHING WORK WELL:
STRONGLY AGREE, AGREE, NEUTRAL, DISAGREE, STRONGLY DISAGREE**

Arthur Best
Professor of Law
University of Denver Sturm College of Law

See Abstract on next page.

**STUDENT EVALUATIONS OF LAW TEACHING WORK WELL:
STRONGLY AGREE, AGREE, NEUTRAL, DISAGREE,
STRONGLY DISAGREE**

ABSTRACT

Academics in the fields of psychology and education generally describe student evaluations of teaching as reliable and useful. On the other hand, law professors often criticize them as unreliable and impaired by students' biases. This Article considers resolving these discrepant views by paying close attention to the various purposes for which student evaluations of teaching are used. For some uses, such as guidance for students in course selection, shortcomings of the evaluations would be of slight consequence. For promotion or tenure decisions, despite law professors' skepticism, schools should use the data to identify outlier instructors. Basing conclusions only on large numerical differences among faculty should protect faculty members from unfair consequences caused by students' biases, since the effects of biases (if present) are likely to be relatively small. It is also consistent with the modern consensus among educational researchers.

The Article also reports findings from analysis of a large number of law school evaluation of teaching forms. Virtually all of them use phraseology that ignores the collaborative nature of teaching and learning. They focus attention on the professor, with the unintended consequence of portraying students as passive participants in their education. The Article recommends revising questionnaires to have a balance between terminology that ignores students' roles and terminology that reflects them. With regard to other attributes, there are large variations among different law schools questionnaires. The Article documents those differences and identifies some that may be problematic.

**STUDENT EVALUATIONS OF LAW TEACHING WORK WELL:
STRONGLY AGREE, AGREE, NEUTRAL, DISAGREE, STRONGLY DISAGREE**

Author's Name on Cover Page¹

I. Introduction

American law students spend more than 30,000 hours each academic year providing evaluations of their courses² If a single person worked forty-hour weeks to do all those evaluations, the task would take at least fifteen years.³ Does this immense amount of work provide anything of value to anyone? Scholars who specialize in the study of teaching and learning conclude

¹ Professor of Law, (author's school shown on preceding cover page). This Article is based on a presentation by the author at the Association of American Law Schools Conference on "New Ideas for Law School Teachers: Teaching Intentionally" in June, 2006, and on law school student evaluation of teaching forms provided to the author by many participants in that conference. This article was supported by a research grant from the Sturm College of Law. Karen Lamprey, Darren McHugh and Carrie Stanley, students at the Sturm College of Law provided thoughtful and very helpful research assistance. I am grateful to Rachel Best, Stephen L. Pepper and Eli Wald for comments on drafts. Errors, of course, are mine.

² The ABA Section of Legal Education's statistical report on law school enrollment states that in academic year 2006-2007, at all of the schools then-accredited by the ABA, enrollment was 148,696. Almost all of those schools conduct student evaluations of teaching. The 30,000 hours figure is derived as follows: assuming that each student evaluates six courses a year (probably a low estimate), and each student takes about two minutes to prepare each evaluation, and assuming that there are about 150,000 students, the minutes devoted to this work would equal 1,800,000 (a number of minutes equal to 30,000 hours).

³ The estimate of fifteen years of a single person's work to accomplish this task gives this poor hypothetical individual two weeks off per year.

generally, but not unanimously, that student evaluations of teaching provide useful information that correlates fairly well with actual learning by students in the courses they evaluate.⁴ On the other hand, many law professors have negative views of student evaluations of teaching.⁵

This Article examines current law school practices related to student evaluations of teaching. Bearing in mind that law school faculty and administrators are probably not qualified to resolve the empirical disputes about the reliability and validity of student evaluations of teaching, and recognizing also that perceptions of illegitimacy deserve respect, this Article concludes that the proper use of student evaluations of teaching depends on the particular purpose the data may serve. Student evaluations may be used in a variety of contexts: 1) assignment of professors to courses, 2) course selection by students, 3) self-improvement by professors, and 4) promotion and tenure decisions by faculty and administrators. They also may, as an unintended consequence of their dissemination, create or reinforce ideas about the roles of students and professors in law school teaching and learning.

For professor assignment and student course selection,

⁴ See Section II, Education Scholars on Student Evaluations of Teaching, *infra*.

⁵ See Section III, Law Professors on Student Evaluations of Teaching, *infra*.

current practices are probably fine. For professors' self-improvement, the timing and style of student evaluations should be improved. For promotion and tenure decisions, data from student evaluations should be used only to identify outliers. Possible biases and users' typical lack of statistical sophistication make it unwise to rely on small numerical differences among scores attained by various professors or by individual professors in various courses or over periods of time as a basis for decisions on matters as important as promotion and tenure.

In connection with the messages that evaluation of teaching questionnaires may provide to students about how students learn and how professors should teach, this Article establishes that almost all evaluation questionnaires incorporate the notion that learning is a passive activity in which professors provide and students receive knowledge. It suggests that questions be revised to orient students more to the collaborative nature of the teaching-learning process.

This Article also provides an empirical examination of student evaluation of teaching forms currently in use. They vary widely in terms of the topics they cover and the styles they use to elicit responses. Users of the data should be aware of differences between global and more specific questionnaire items, and of some variation in reliability between those two

types of items. The variation among forms also suggests that schools might benefit from awareness of techniques and inquiries being used at other institutions. For individuals who are disposed to treat the evaluation process with skepticism, the lack of apparent consensus on how to conduct evaluations might reinforce their belief that any particular school's process is not valid.

II. Education Scholars on Student Evaluations of Teaching

Thousands of articles have been published about student evaluations of teaching. By 1990 one scholar in the field could report that his own collection of books and articles on evaluating instruction contained more than 2,000 items.⁶ Many articles have reviewed portions of this literature,⁷ and there have been review articles reviewing review articles.⁸ In 1987 H.

⁶ K. A. Feldman, "Instructional Evaluation," 4 *The Teaching Professor*, 5-7 (1990).

⁷ See, e.g., L. M. Aleamoni, *Typical faculty concerns about student evaluation of teaching*, L. M. Aleamoni (Ed.), *Techniques for evaluation and improving instruction. New directions for teaching and learning* (1987), R. A. Arreola, R. A. (1995). *Developing a comprehensive faculty evaluation system: A handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system* (1995), F. Costin, W. T. Greenough, and R. J. Menges, *Student ratings of college teaching: Reliability, validity, and usefulness*, 41 *Review of Educational Research* 511 (1971), W. J. McKeachie, *Student ratings of faculty: A Reprise*, 65 *Academe* 384-397 (1979).

⁸ See, e.g., H. G. Murray, *Evaluating University Teaching: A Review of Research* (1980); W. E. Cashin, *Student Ratings of*

W. Marsh published a detailed review of then-existing research concluded that student evaluations of teaching “are clearly multidimensional, quite reliable, reasonably valid, relatively uncontaminated by many variables often seen as sources of potential bias.”⁹

A summary of relevant research published by Kenneth A. Feldman in 1997 concluded that many faculty members had come to believe certain myths and half-truths as accurate descriptions of student evaluations of teaching. On the basis of a massive review of literature, it characterized the following ideas as *myths*: students are not capable of making consistent judgments about teaching, student ratings are popularity contests that favor warm and humorous instructors, and ratings are both unreliable and invalid.¹⁰ This same summary also reported that gender of instructors does not bias students’ ratings in any degree that would be significant in practical terms, although some studies show some statistical differences that favor women over men.¹¹

Teaching: The Research Revisited (IDEA Paper No. 32) (1995).

⁹ H. W. Marsh, *Students’ Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research*, 11 *Int’l Journal of Educational Research* 253 (1987).

¹⁰ Kenneth A. Feldman, *Identifying Exemplary Teachers and Teaching: Evidence from Student Ratings*, reprinted in R. P. Perry and J. C. Smart, *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, 2007.

¹¹ Feldman, *supra*, at 96-97. See also, Kenneth A. Feldman,

A 2001 article by Michael Theall and Jennifer Franklin also based on a large body of prior research, concluded that "student ratings are generally valid and reliable."¹² The article reviewed prior work examining, among other topics, whether students are unqualified to rate instructors, whether ratings measure mainly "popularity," whether ratings are related to learning, and whether gender biases affect ratings. With regard to each of these possible flaws in student evaluations of teaching, the authors concluded that the consensus is that there is little proof that any of them exist or are significant.¹³

Michael and Kenneth A. Feldman published a 2007 update of Feldman's 1997 literature review; they did not report much change in the basic understanding of student evaluations of

College Students' Views of Male and Female College Teachers: Part II - Evidence from Students' Evaluations of their Classroom Teachers, 34 *Research in Higher Education* 151, 176 ("[T]he majority of studies did not show a statistically significant difference between the two genders... however, when statistically significant differences were found, more of them favored women than favored men") (1993); Anthony G. Greenwald, *Validity Concerns and Usefulness of Student Ratings of Instruction*, 52 *American Psychologist* 1182 (1997) (summarizing four related articles by Marsh and Roche, d'Apollonia & Abrami, Greenwald & Gillmore, and McKeachie that conclude that "ratings are relatively unaffected by potential biases," "there is little evidence of bias in ratings," "ratings measures show moderate correlations with achievement..." and "student ratings provide valid, albeit imperfect, measures of teaching effectiveness" (respectively)).

¹² Michael Theall and Jennifer Franklin, *Looking for Bias in All the Wrong Places: A Search for Truth or a Witch Hunt in Student Ratings of Instruction?* 109 *New Directions for Institutional Research* 45 (2001).

¹³ Theall and Franklin, *supra*, at 48-50.

teaching.¹⁴ They did find, though, that researchers and practitioners in the field had continued to be active and that efforts to relate their research to practice had intensified. Another new finding was that numerous articles have been published outside the usual discipline of the academic study of education, and that most of these articles have sought to establish that student evaluations of teaching are biased, cause grade inflation, threaten academic freedom, and harm the general quality of higher education. These newer articles were described as typically based on methodologies that vary substantially from standard practice and as “opinion pieces.”¹⁵

Another leading scholar in this field summarized the state of research in 2007:

Numerous studies have shown SETs [student evaluations of teaching effectiveness] to be valid in relation to a variety of outcome measures broadly accepted by classroom teachers (e.g., learning inferred from classroom and standardized tests, student motivation, plans to pursue and apply the subject, positive affect, experimental manipulations of specific components of teaching, ratings by former students, classroom observations by trained external observers, and even teacher self-evaluations of their own teaching effectiveness...) ¹⁶

¹⁴ Michael Theall and Kenneth A. Feldman, Commentary and Update on Feldman’s (1997) “Identifying Exemplary Teachers and Teaching: Evidence from Student Ratings” in R. P. Perry and J. C. Smart, *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, 2007.

¹⁵ Theall and Feldman, *supra* note 14, at 134.

¹⁶ Herbert W. Marsh, Do University Teachers Become More Effective With Experience? A Multilevel Growth Model of Students’ Evaluations of Teaching Over 13 Years, 99 *J. Educational Psychology* 775, 775 (2007) (references omitted).

This statement represents the mainstream consensus among education researchers concerning the usefulness of student evaluations of teaching.

None of these reviews consider racial bias. The lack of a body of research on racial bias in connection with student evaluations is troubling. Reports of personal experiences suggest that racial prejudice may affect evaluations,¹⁷ but consideration of this serious possibility has been minuscule in the array of research on the general topic of student evaluations.¹⁸ The findings on gender bias, that evidence shows

See also Robert J. Menges and Ann E. Austin, *Teaching in Higher Education, Handbook of Research on Teaching*, 4th Ed., 2001, page 1144 (“[T]he body of research supporting student evaluations is far more substantial than for any other component of faculty evaluation...Often people are skeptical about student evaluations (a) because they fail to understand or acknowledge the worth of that research; (b) because users overinterpret quantitative data (assuming that there are meaningful distinctions in small statistical differences) ...”)

¹⁷ See, e.g., Pamela J. Smith, *Teaching the Retrenchment Generation: When Sapphire Meets Socrates at the Intersection of Race, Gender, and Authority*, 6 *Wm. & Mary J. Women & Law*. 53, 167 (describing racial stereotypes in evaluations),

¹⁸ Small sample sizes and other methodological problems have hampered efforts to study this problem. See, e.g., Deborah L. Brake, *Perceiving Subtle Sexism: Mapping the Social-Psychological Forces and Legal Narratives that Obscure Gender Bias*, 16 *Colum. J. Gender & Law* 679, 721 (2007) (“Part of the difficulty in challenging the use of student evaluations is that it is difficult to determine whether sex or race bias has influenced student evaluations in any particular case”); Therese A. Huston, *Pedagogy and Social Justice: Race and Gender Bias in Higher Education: Could Faculty Course Evaluations Impede Further Progress Toward Parity?* 4 *Seattle J. Soc. Just.* 591, 598

either no effect or small effects, might be suggestive of the scope of possible racial prejudice effects, but that is an empirical question to be answered in future research.

III. Law Professors on Student Evaluations of Teaching

Many law professors have written about student evaluations.¹⁹ The topic, naturally, has great personal interest

(2006) (stating that several studies report that student evaluations of teaching leave faculty of color at a disadvantage but supporting this statement mainly with reference to a single unpublished study); Kathryn Pourmand Nordick, *Essay, A Critical Look at Student Resistance to Non-Traditional Law School Professors*, 27 *W. New Eng. L. Rev.* 173, 188, 191 (2005) (finding statements that indicated bias in a random sampling of student evaluations); Michael A. McPherson and R. Todd Jewell, *Leveling the Playing Field: Should Student Evaluation Scores be Adjusted?* 88 *Social Science Quarterly* 868, 875 and 875, note 10 (2007) (stating in text that "Race appears to play a significant role in SET [student evaluation of teaching] scores in our data, with white instructors earning ... higher scores than their nonwhite colleagues" but explaining in a footnote that "This conclusion should be viewed with some skepticism as white instructors teach the majority of courses in our sample and 19 of the 22 instructors are white."); Roger W. Reinsch, Susan M. Des Rosiers and Amy B. Hietapelto, *Evidentiary And Constitutional Due Process Constraints on The Uses By Colleges and Universities of Student Evaluations*, 32 *J. College and University Law* 75, 84 (2005) (stating in text that "research shows many examples of biases," but supporting that assertion primarily with description of a homophobic statement made by two students in a class).

¹⁹ See articles discussed in this section. See also: Judith D. Fischer, *How to Improve Student Ratings in Legal Writing Courses: Views from the Trenches*, 34 *U. Balt. L. Rev.* 199 (2004); Daniel Gordon, *Does Law Teaching Have Meaning? Teaching Effectiveness, Gauging Alumni Competence, and the MacCrate Report*, 25 *Fordham Urb. L.J.* 43 (1997); Angela P. Harris, *"A(Nother) Critique of Pure Reason": Toward Civic Virtue in Legal Education*. 45 *Stan. L. Rev.* 1773 (1993); Gerald F. Hess,

to professors who are evaluated, and it often seems easy to collect data on the subject.²⁰ A number of articles discuss the student evaluation of teaching process as one aspect of the hardships that many professors of color have faced as their presence on law faculties has increased.²¹

Improving Teaching And Learning In Law School: Faculty Development Research, Principles, And Programs, 12 Widener L. Rev. 443 (2006); Melissa J. Marlow, Blessed Are They Who Teach An Upper-Level Course, For They Shall Earn Higher Student Ratings, 7 Fla. Coastal L. Rev. 553 (2006); Mary Pat Treuthart, A Perspective On Teaching And Learning Family Law, 75 UMKC L. Rev. 1047 (2007); John Mixon, Continuous Quality Improvement, Law, and Legal Education, 43 Emory L. J. 393 (1994).

²⁰ For a related observation, see Clarence Morris and C. Robert Morris, Jr., *Morris on Torts*, 2d Ed., 1980 at 2-3 ("A lawyer who defends bottlers of soft drinks against claims for injuries alleged to have come from foreign matter in their product said long ago (when the price, as well as the product, both were more modest than they now are) that he never recommended a settlement. 'After all,' he said, 'all a man needs to go into business against us is a bug and a nickel.'") If a professor is aggrieved about student evaluations, all he or she needs to produce an article is a group of student evaluations and some other information about the course or courses the students evaluated; these materials are readily available to almost everyone in academe.

²¹ See, e.g., Okianer Christian Dark, Just My 'Magination, 10 Harv. BlackLetter L.J. 21 (1993); Richard Delgado & Derrick Bell, Minority Law Professors' Lives: The Bell-Delgado Survey, 24 Harv. C.R.-C.L. L. Rev. 349 (1989); Trina Grillo, Tenure and Minority Women Law Professors: Separating the Strands, 31 U.S.F. L. Rev. 747 (1997); Reginald Leamon Robinson, Teaching from the Margins: Race as a Pedagogical Sub-Text, 19 W. New Eng. L. Rev. 151 (1997); Pamela J. Smith, Teaching the Retrenchment Generation: When Sapphire Meets Socrates at the Intersection of Race, Gender, and Authority, 6 Wm. & Mary J. Women & L. 53 (1999), Pamela J. Smith, The Tyrannies of Silence of the Untenured Professors of Color, 33 U.C. Davis L. Rev. 1105 (2000); and Donna E. Young, Two Steps Removed: The Paradox of Diversity Discourse for Women of Color in Law Teaching, 11 Berkeley Women's L.J. 270 (1996).

Some law professors have taken positions that are consistent with those of mainstream education scholars. For example, Ronald H. Silverman argues that students are well-positioned to evaluate teaching,²² since they experience a range of teaching styles, can make comparisons among professors even if they do not have “perfected definitions of effective law teaching,”²³ and “know more about their own boredom and confusions that most monitoring faculty members and deans.”²⁴

Overall, though, in articles by law professors about student evaluations, most attention is directed towards shortcomings of that process. William Roth reported in 1984 on a survey of student evaluation of teaching practices of American law schools. He asserted that “no one really believes that the process does much good,”²⁵ that examination of evaluation forms from 156 schools showed “shocking”²⁶ differences among them, and that for administrative purposes “law school teacher evaluations are not taken very seriously,”²⁷

In 1990, Richard Abel published a heartfelt and frequently-cited article seeking “to document error and bias ... to

²² Ronald H. Silverman, *Weak Law Teaching, Adam Smith and a New Model of Merit Pay*, 9 *Cornell J. L. & Pub. Policy* 267 (1999-2000).

²³ Silverman, *supra* note 22, at 361, 390-391.

²⁴ Silverman, *supra* note 22, at 361, 390-391.

²⁵ William Roth, *Student Evaluation of Law Teaching*, 17 *Akron L. Rev.* 609 (1984).

²⁶ Roth, *supra*, note 25, at 610.

²⁷ Roth, *supra*, note 25, at 609.

understand how they enter the evaluation process and to consider how we might reduce them.”²⁸ He noted, based on survey data, that almost all law schools conduct student evaluations of teaching with questionnaires administered at the end of each course but that “almost nobody is denied tenure because of teaching.”²⁹ Based on a review of many empirical studies, Professor Abel explored the effects inappropriate variables and methodological flaws. He concluded that student evaluations of teaching should “remain an important tool”³⁰ but that students should give narrative responses rather than numerical ones, that all students be required to evaluate each course, that students “should be encouraged to reveal their political agreements and disagreements” and that the process should elicit “the hidden biases toward women, racial minorities and political dissidents” so that users of the evaluations could control for those factors.

A later article by Daniel Gordon³¹ argues that student evaluations of teaching measure only “student consumer attitudes” or “technical teaching proficiency”³² and that they

²⁸ Richard L. Abel, *Evaluating Evaluations: How Should Law Schools Judge Teaching?*, 40 *J. Legal Ed.* 407 (1990).

²⁹ Abel, *supra*, note 28, at 417.

³⁰ Abel, *supra*, note 28, at 452.

³¹ Daniel Gordon, “Does Law Teaching Have Meaning? Teaching Effectiveness, Gauging Alumni Competence, and the MacCrate Report,” 25 *Fordham Urb. L. J.* 43 (1997)

³² Gordon, *supra*, note 31, at 45.

“possess little worth in gauging and improving teaching effectiveness.”³³ The Article supports this conclusion by describing the author’s experience of having taught two sections of a course one semester. He implies that differences in ratings in the two sections would point to shortcomings in the evaluation process.

The only evidence to support the idea that the teaching was the same in each section is the author’s own self-assessment. He reports that he received “excellent” evaluations in one section and “mediocre” evaluations in the other section.³⁴ The author reports that the overall average of his ratings on a five-point scale was 4.2 in one section and was 3.7 in the other. He also reports specific differences on certain questionnaire items, such as average scores of 4.7 and 5.0 for the sections’ responses to a question about instructor preparedness. These figures are not put in the context of the number of responses for any questions by students in either section, which prevents an analysis of the possibility that the differences were due to random chance.

Professor Deborah Maranville provides a discussion of student evaluations at a school where, early in her career, she was denied tenure with student evaluations being “the stated

³³ Gordon, *supra*, note 31, at 51.

³⁴ Gordon, *supra*, note 31, at 51.

reason" for that denial.³⁵ Aware that the consensus of research on student evaluations is that gender bias may or may not exist, but that if it does exist it likely introduces only small distortions into evaluations, she notes that "evidence" concerning gender bias in law school student evaluations "is primarily anecdotal in nature,"³⁶ but suggests that she was at risk for gender-biased evaluations. However, she also describes her teaching as having had numerous shortcomings including deficiencies in energy and enthusiasm.³⁷

A recent article by Professor Deborah J. Merritt focuses on a detailed and comprehensive description of psychological research about nonverbal factors. It argues that there is a link between nonverbal behavior and teaching evaluations that "*hints* that conventional evaluations bear little relationship to student learning."³⁸ On that basis, and to protect professors of color from being affected by students' prejudices that may be

³⁵ Deborah Maranville, "Classroom Incivilities, Gender, Authenticity and Orthodoxy, and the Limits of Hard Work: Four Lenses for Interpreting a "Failed" Teaching Experience," 12 Wm. And Mary J. Women and Law 699, 717 (2006).

³⁶ Maranville, *supra*, note 35, at 720.

³⁷ Maranville, *supra*, note 35, at 711.

³⁸ Deborah J. Merritt, *Bias, the Brain, and Student Evaluations of Teaching*, 82 St. John's Law Review 235, 270 (2008) (emphasis added). The limitation indicated by the word "*hints*" contrasts with another statement in the Article, "[e]xtensive research by psychologists and educators convincingly demonstrates that [student evaluations of teaching] are biased" (at page 238), which conveys a message different from the professional consensus that some small biases may have been established in some studies of the subject.

reflected in conventional evaluation processes, the Article concludes that law school should replace “conventional student evaluations with more meaningful processes.”³⁹ The article recommends student evaluation systems that encourage reflective thought and particularly endorses the suggestion of Gregory Munro that a facilitator meet with small groups of students to “gather their impressions of a course and instructor.”⁴⁰

These brief descriptions of some law professors’ scholarship on student evaluations of teaching show an overriding sense of dissatisfaction with the process and a good deal of skepticism about whether it is worthwhile at all. Roth and Abel each conclude that the data are not used in any significant way for tenure decisions, but they are nonetheless critical either of the way in which it is collected or the ways in which it might otherwise be used. Abel, Maranville and Merritt identify different methodological flaws in typical practices, and Gordon’s analysis may illustrate the hazards of making numerical comparisons in the absence of statistical controls. These views from scholars whose primary work is outside the fields of psychology and the study of teaching and learning differ strongly from most of the findings of

³⁹ Merritt, *supra*, note 38, at 285.

⁴⁰ Merritt, *supra*, note 38, at 281.

researchers in those other fields.⁴¹

III. Sensible Use of Student Evaluations as Indicators of Teaching Quality

Suggested General Analysis. What use of student evaluations of teaching is fair and sensible for: 1) assignment of professors to courses, 2) course selection by students, 3) self-improvement by professors, and 4) promotion and tenure decisions by faculty and administrators? Many law professors have argued that the current student evaluation of teaching process is seriously flawed. Regardless of whether that concern actually reflects a true understanding of how well student evaluations

⁴¹ It would be surprising for law professors to know a narrow field like student evaluations of teaching as well as professors in education and psychology would, considering that some of those professors devote their entire careers to that topic. Certainly this Article is itself an example of an outsider's effort to bring a careful but necessarily summary reading of a body of work to a non-expert audience. Another example of these differences in depth might be Professor Merritt's non-critical references to the "Dr. Fox" experiment as establishing that expressive professors receive higher evaluations than non-expressive professors regardless of the quality of students' learning (Merritt, *supra*, note 38, at 242-243) with virtually no description of a large body of scholarship that questions most of the study's conclusions. For a small selection of that literature, see Herbert W. Marsh and Michael J. Dunkin, *Students' Evaluations of University Teaching: A Multidimensional Perspective*, 241, 299-302, in Raymond P. Perry and John C. Smart, *Effective Teaching in Higher Education: Research and Practice* (1997) (stating that the authors of the Dr. Fox study and its critics agree that the study was methodologically weak, and referring to more than a dozen articles evaluating it.)

may work, the persistence of that view is a fact which members of an academic community must acknowledge.⁴² If techniques for using student evaluation of teaching data can be designed to be respectful of that skepticism, adopting them would increase the perceived legitimacy of decisions to which those data contribute.

Considering the controversy about student evaluations of teaching, deciding how to use them for any purpose should balance two considerations. The first factor is whether the function involves judgments or discriminations that can logically be made only with a high level of detailed and precise information. The second factor is the size of the risk associated with implementing a decision on imperfect information.

Assigning Instructors to Courses. How useful are student evaluations of teaching for the administrative function of assigning professors to courses and, in schools with multiple sections of first-year students, establishing the lineup of instructors for each first-year section? This task does not require much detailed information, since the options are typically few. A general impression of teaching styles and a

⁴²It may also be that differences among various schools' questionnaires (illustrated in the next section of this Article) could contribute to skepticism about their reliability, and might raise doubts about whether any particular form has an optimum selection of topics and types of items.

general sense of an instructor's relative performance in large-enrollment and small-enrollment courses would probably be sufficient to provide some useful information relevant to this function. Using information from student evaluation forms only to identify outliers - instances where the array of evaluations seems extremely better or extremely worse than typical evaluations given by a school's students - could make a sensible contribution to this analysis, bearing in mind that the analysis does not require detailed information about particular aspects of an instructor's teaching. Broad and qualitative impressions are probably useful, and the shortcomings of typical evaluation forms do not prevent their users from deriving impressions of that level of generality.

For course assignments, the costs of a "wrong" decision are certainly low. These decisions almost always involve tradeoffs and predictions of faculty and student preferences. Comparing any two possible decisions would likely suggest that either one had strengths and weaknesses. In sum, pairing up instructors and courses is a function that is important but that has very low risks of significant harm even if done poorly.

Students' Course Selection. How helpful can student evaluations of teaching be to students who would like to select elective courses that will be the most rewarding for them? Again, the need for precision is low and the costs of "error"

are low. If a student uses the data impressionistically, mainly to identify outliers - professors whose evaluations in a particular course seems far outside the typical range of evaluations students give to faculty at a particular school - the data are likely to be helpful.

Professors' Self-Improvement. Typically, student evaluation of teaching uses a standard form at the end of each semester. The process is thus generalized and infrequent. To be most useful for an instructor's self-improvement, feedback would be specific and frequent. Thus, current practice involves a large gap between the ideal of specific and frequent and the practice of general and rare.

Forms currently in use may be of some value to professors, again if the data are analyzed with a view towards identifying outlier instances. Doubts about the reliability of student evaluations of teaching are unimportant in the context of professors' self-improvement, since faculty members can supplement their institutions' data collection in a variety of ways. Faculty members can solicit anonymous written responses to daily inquiries, can organize focus groups or quality circles,⁴³ and can invite colleagues to visit or watch videos of their

⁴³ Gerald F. Hess, *Student Involvement in Improving Law Teaching and Learning*, 67 U.M.K.C. L. Rev. 343 (1998); Eric W. Orts, *Quality Circles in Law Teaching*, 47 J. Legal Educ. 425, 425-26 (1997) (describing "quality circles").

classes for the purpose of informal discussions of approaches and outcomes.

Institutional Decisions on Promotion and Tenure. Promotion and tenure decisions are of vital importance to institutions and to faculty members. When an institution grants tenure, it is committing resources for a period of time often measured in decades. When a faculty member is denied tenure, that person's investment of time and labor at the institution loses most of its anticipated value to the faculty member, and the faculty member's opportunities to continue in teaching are decreased. The stakes are high, and interest in developing a fair and accurate assessment of a candidate's teaching ought also to be high.

In light of the controversy about whether student evaluations of teaching provide useful information, two conclusions seem justified. One is that because of substantial support in the academic literature for the validity and reliability of these evaluations, they should receive *some* consideration.⁴⁴ The second conclusion is that the consideration

⁴⁴ Ideally, an institution should provide a clear statement of the ways in which it will use student evaluation data for promotion and tenure decisions. It may be common, however, for schools to be vague about this. See, e.g., Daniel Kating, *A Comprehensive Approach to Orientation and Mentoring for New Faculty*, 46 *J. Legal Ed.* 59, 65-66 (1996) (each voter in the tenure process can decide personally how to use student evaluation data at the author's school; a description of the

they receive should be constrained.

When an instructor's student evaluations show scores that are much worse than typical scores at an institution, or when the narrative responses to open-ended queries seem heavily weighted towards negative reactions, those results should lead an institution to be extremely cautious with regard to concluding that the instructor meets its standards for promotion or tenure. Students' evaluations should naturally be just one of many components of a full evaluation of an instructor's performance.⁴⁵ But against the background of significant (though controverted) scholarly conclusions that student evaluations of

process provided to junior colleagues stated that numerical evaluations are considered in light of narrative comments).

Courts have supported institutions' use of student evaluations in personnel decisions. See, e.g., *Agarwal v. Regents of the University of Minnesota*, 788 F.2d 504 (8th Cir. 1986) (poor student evaluations were among factors that supported a finding of good cause for termination of employment), *Brousard-Norcross v. Augustana College Assn.*, 935 F.2d 974, 976 (8th Cir. 1991) (negative student evaluations are a legitimate factor for tenure evaluation). These cases are cited in John D. Copeland and John W. Murray, Jr., *Getting Tossed From the Ivory Tower: The Legal Implications of Evaluating Faculty Performance*, 61 *Mo. L. Rev.* 233 (1996).

⁴⁵ See, e.g., Herbert W. Marsh and Michael J. Dunkin, *Students' Evaluations of University Teaching: A Multidimensional Perspective*, 241, 311, in Raymond P. Perry and John C. Smart, *Effective Teaching in Higher Education: Research and Practice* (1997) (student evaluations of teaching, while supported by more research than any other methods of evaluating teaching, should be used in conjunction with other inputs, particularly for tenure or promotion decisions).

teaching are reliable and valid, using them for the purposes of gross rather than fine distinctions seems legitimate.⁴⁶

The possibility that improper biases can lead to lower evaluations of some instructors should lead those who use the data to ignore small variations among instructors. But to disregard student evaluations entirely because of the risk of improper bias would be to act as though improper influences are likely to have large effects on overall evaluations. For gender biases, studies have shown only small effects. For racial biases, studies have not shown such effects, although the number of studies on this important topic is small.

There is another reason why student evaluations of teaching should be used only for a rough sorting function rather than for precise discriminations. Most of the law school personnel who might be tempted to use evaluation scores for subtle distinctions are probably ignorant of the statistical legitimacy of that use of the data. Where one instructor's overall average score might be 3.0 on a five point scale and another instructor's overall average might be 3.5, statisticians can deal with the question of whether the difference between those

⁴⁶W. J. McKeachie, *Good Teaching Makes a Difference - And We Know What It Is*, 396, 402, in Raymond P. Perry and John C. Smart, *Effective Teaching in Higher Education: Research and Practice* (1997) (distinguishing among faculty based on numerical differences can mask the likely circumstance that each of the faculty members being compared is a good teacher).

two numbers is statistically significant - that is, whether it is unlikely to have resulted from chance alone. They would take into account the size of the difference, the number of respondents, and the degree to which the responses were bunched or spread out. Looking at a difference between two numbers without taking these factors into account fails to account for the likelihood that the difference was due to random chance.

Another issue that can lead to overvaluing the summary data from evaluation forms is the use of averages that assign numerical values to concepts such as "agree" and "strongly agree." These scales assume that there is a clear understanding of a linear relationship in strength between the scales' terms. Table 1 provides an example: in a scale with 1 representing the worst performance and 5 representing the best performance, an instructor's average score on a questionnaire item with 24 respondents would be 3.0 if the instructor received any of the following combinations of scores: [Table 1 is on next page.]

Table 1: Groups of 24 Ratings on 5-Point Scale
 That Each Produce an Average of 3.0
 (with * indicating a single rating in each group)

Rating and Score	Group 1	Group 2	Group 3	Group 4
"Strongly Disagree" (1)	***** *****	*****	*** *** ***	
"Disagree" (2)		*****		
"Neutral" (3)		*****	*****	***** ***** ***** *****
"Agree" 4		*****		
"Strongly Agree" (5)	***** *****	*****	*** *** ***	

The point system is entirely arbitrary. The difference between "Disagree" and "Strongly Disagree" is one point. The difference between "Agree" and "Neutral" is also one point. No one can know whether the differences in intensity that a particular respondent associates with the terms "disagree," "strongly disagree" and "neutral" are the same or different.

Also, the Table shows that very different arrays of assessments can produce the same average rating. For that reason, paying extremely close attention to average numbers produced by student evaluations is risky.

Finally, some research has shown that for summative decisions about teaching effectiveness, global items work better than more specific items about particular aspects of teaching, and that validity of global measures is higher than validity of more specific items and less subject to variations caused by changes in instructional settings.⁴⁷ This suggests that if a school's student evaluation form uses many narrow-focus items, special care in using the data such forms produce.

⁴⁷ Philip C. Abrami, Sylvia d'Apollonia, and Steven Rosenfield, *The Dimensionality of Student Ratings of Instruction: What We Know and What We Do Not*, 321, 357, in Raymond P. Perry and John C. Smart, *Effective Teaching in Higher Education: Research and Practice* (1997).

IV. Recognizing and Assessing Student Evaluation Questionnaires' Implicit Messages about Teaching and Learning

The forms law schools use for student evaluations of teaching necessarily convey messages about what aspects of teaching schools consider important and about how those aspects might be measured. Perhaps unintentionally, the forms studied for this Article⁴⁸ overwhelmingly convey the idea that learning is a passive activity and that teaching consists of a one-way delivery process, with information and skills directed to students by the professor. Rarely represented is an alternative view of the teaching-learning process, that it is a collaborative enterprise with work to be done by both instructors and students.⁴⁹

Many forms use items similar to "How successful was the instructor in arousing your intellectual interest and excitement?" An alternate approach to the same topic would be "Did you become intellectually interested and excited by this course?" The first style of questionnaire item explicitly puts the responsibility for interest and excitement on the professor. The second style focuses on the student's experience.

⁴⁸ Section V, *infra*, and the Appendix describe this study's sampling methodology.

⁴⁹ See, e.g., John D. Bransford, Ann L. Brown, and Rodney R. Cocking, *How People Learn: Brain, Mind, Experience, and School* (1999) (compared with passive learners, active learners are better able to understand and communicate difficult concepts).

Many forms ask questions about specific student conduct, probing the student's typical level of preparation for classes and the student's attendance record. For inquiries that relate to the effectiveness of teaching, questions can similarly frame the topic to recognize the student's participation in the enterprise or can frame the topic in a way that highlights the professor's individual conduct. Table 2 offers examples of each type of framing. Its classification of these questions naturally involves some debatable judgments. But overall, the intent of the table is to compare questions that identify and respect a student's own participation in the teaching-learning process with questions that ask a student to comment explicitly and only on the conduct of the professor. [Table 2 is on next page.]

Table 2: Questionnaire Items on Selected Topics
Focusing on Student's or Professor's Role

Topic	Professor's Role	Student's Role
Learning	How successful was the instructor in imparting substantive knowledge of legal rules? ⁵⁰	I learned a great deal in this course. ⁵¹
Organization	The course was well prepared and organized. ⁵²	I understood the professor's organization of the course. ⁵³
Pace	Was the course so organized that the goals were clear, and was the pace and emphasis of instruction appropriate? ⁵⁴	I was able to keep up with the workload. ⁵⁵
Rapport	Professor's relationship with students. ⁵⁶	I was comfortable asking questions in this course. ⁵⁷

⁵⁰ Creighton U

⁵¹ University of Utah

⁵² American University.

⁵³ Drake.

⁵⁴ Washburn

⁵⁵ Denver.

⁵⁶ University of Nebraska

⁵⁷ University of Utah

Applying the style of categorization represented in this table to all of the questions on the forms analyzed for this study, out of a total of 656 questions on all of the forms, only about 14 questions recognize the student's role in teaching and learning. Even if the characterizations of some questions might be questioned, this analysis shows that the overwhelming approach in these forms is to focus attention solely on the work of the professor while ignoring the desirability of active engagement by students.

In considering student evaluation forms, the usual issue is whether the questionnaire items are likely to *elicit* accurate information *from* students. In contrast, for analyzing how questionnaire items may depict the roles of students and professors in law school teaching and learning, the issue is whether they may *convey* false information *to* students. Almost all of the questionnaire items law schools use focus on the professor's conduct and ignore the student's own necessary engagement in the process of learning. Probably good students understand that being deeply engaged with a course contributes to their learning, but it seems unfortunate that almost all of the questions in the forms convey the implicit idea that a student is a passive audience member.

A comparison illustrates this point. One form asks "How successful was the instructor in imparting substantive knowledge

of legal rules?" Another form's parallel item is "I learned a great deal in this course." Obviously a student's degree of learning in a course will be connected to attributes of the instructor. But using phraseology that ignores the student's role presents a false view of the reality of the teaching-learning process. Whether propagating this false view actually causes any harm is an empirical question suitable for future research. However, acknowledging the risk that this may occur is easy to do, and rewriting some questions so that they lead students to focus on "I..." statements rather than on "The professor..." statements might help orient students to the importance of their own contributions to their learning.

This inquiry is not intended to fault professor-centric questions in terms of their ability to develop worthwhile information about professors' attributes or about professors' contributions to the teaching-learning process. Rather, it shows that questions typically used to gather this information may establish or reinforce assumptions that instructors have total responsibility for students' learning. It is likely possible to revise current evaluation forms so that they could collect the data they seek with a mix of two kinds of questionnaire items: some that continue the focus on professors and some that are free from the current typical implicit message that students are passive consumers of the work of professors.

V. Current Law School Practices

Methodology and General Findings. Student evaluation of teaching forms from thirty-eight law schools were given to the author by attendees at an Association of American Law Schools conference.⁵⁸ The schools are listed in the Appendix. This sample is not random, but it is large, considering that only 195 law schools are currently accredited by the American Bar Association to confer the J.D. degree.⁵⁹ Referring to the U.S. News and World Report rankings of law schools,⁶⁰ the schools represented in this study come from all of the U.S. News ranking categories ("Top 100," "Tier 3" and "Tier 4").⁶¹ Eighteen are from the publication's "Top 100" (they are about half of the

⁵⁸ Association of American Law Schools Conference on "New Ideas for Law School Teachers: Teaching Intentionally" in June, 2006.

⁵⁹

<http://www.abanet.org/legaled/approvedlawschools/approved.html> (last visited February 8, 2008) states that there are 195 institutions currently accredited to confer the J.D. degree. One other law school, the U.S. Army Judge Advocate General's School, is also accredited but does not confer the J.D. degree.

⁶⁰ Using that ranking to describe the variety of schools in this study's sample is not intended to endorse that ranking as accurate. On that issue, see Statement Regarding Law School Rankings, issued by the Association of American Law Schools, the American Bar Association Section of Legal Education and Admissions to the Bar, the Law School Admission Council and the National Association of Law Placement (1990) ("any ranking or rating of law schools, based upon the data the magazine has asked deans to provide, must be meaningless or grossly misleading").http://www.aals.org/about_handbook_sgp_ran.php (last visited February 22, 2008).

⁶¹ See Appendix.

schools in this study, and the "Top 100" are about half of American law schools). Eleven are from the publication's "Tier 3" and nine are from the publication's "Tier 4." The sample's distribution of schools is approximately proportional to the distribution of schools in the U.S. News array. The schools are from all geographic regions and include small, large, public and private schools.

A more comprehensive empirical study of law school evaluation forms was reported in 1987. It showed that virtually all law schools conducted evaluations. The current study does not provide a basis for knowing precisely how many law schools currently conduct these evaluations, but it is likely that most schools still use this process (attendees at the 2006 conference, from many schools, generally agreed that the practice is widespread).

Current law school practices for student evaluation of teaching show wide variation. Schools' evaluation forms differ with regard to almost every characteristic, including obvious ones such as total number of questions and other attributes that may be less apparent such as whether questions focus on the role of students or the role of professors. This lack of uniformity could possibly contribute to the low repute the evaluation process has among some faculty, on the assumption that standard practices ought to have developed for an activity that has been

carried out for decades.⁶² It should be noted, though, that a body of research contradicts such an assumption; that research supporting the accuracy of student evaluations of teaching is comprised of studies that assess evaluation forms with many different attributes.⁶³

As noted in the prior section, some research has shown that global questionnaire items work better for summative decisions about teaching effectiveness than do items that focus on specific aspects of teaching.⁶⁴ This does not mean that questionnaires with many narrow-focus items should be rejected, but it should make their users pay special attention to the reasons for using many detailed questions and also to the method in which they use the data from highly specific questions.

Seeing the range of content and style in current forms could inspire changes by some schools in their questionnaire items. A current form may produce useful data, but thoughtful

⁶²An alternate interpretation would be that many different types of forms can accomplish the basic function, and the differences among forms may not matter much, particularly if users of the data do not use it to make fine or subtle distinctions.

⁶³ Herbert W. Marsh and Michael J. Dunkin, *Students' Evaluations of University Teaching: A Multidimensional Perspective*, 241, 268, in Raymond P. Perry and John C. Smart, *Effective Teaching in Higher Education: Research and Practice* (1997) (research on student evaluations of teaching has been carried out on many different questionnaires, with no uniform set of evaluation components).

⁶⁴ See text, *supra*, at note 47.

consideration of its attributes could lead to changes that would gather information on additional topics or would protect against the risk of overweighing numerical results.

Topics of Questions. Two education scholars have proposed that student evaluations of teaching should cover nine specific attributes of an instructor and a course.⁶⁵ These attributes are: 1) learning or value, 2) instructor enthusiasm, 3) organization or clarity, 4) group interaction, 5) individual rapport, 6) breadth of coverage, 7) examinations or grading, 8) assignments or readings, and 9) workload or difficulty.⁶⁶ While the language used in the law school forms shows great variety, most of the forms cover most of these attributes. However, some of them were ignored in many of the thirty-eight forms studied. Table 3 shows that large numbers of forms fail to give explicit attention to breadth of coverage, exams or workload difficulty. [Table 3 is on next page.]

⁶⁵ Robert J. Menges, Ann E. Austin, "Teaching in Higher Education," *Handbook of Research on Teaching*, Ed. Virginia Richardson, American Educational Research Association, 1143 (2001)

⁶⁶ *Id.*

Table 3: Inquiry Topics
 (percentages of schools with items on each attribute)

Attribute	Schools with Items Related to Attribute
Learning or value of course	100%
Organization or clarity	97%
Instructor enthusiasm	82%
Group interaction	82%
Assignments or readings	68%
Individual rapport	53%
Breadth of coverage	32%
Workload or difficulty	26%
Examinations or grading	24%

Every form includes at least one item related to learning or value of the course. But depending on the use to which the evaluations may be put, and the degree of specificity that users may explicitly or implicitly apply when they use the data, it may be problematic that many forms ignore aspects of the teaching-learning experience that would generally be considered important, such as coverage, assessments and degree of difficulty. If a school's form is relatively brief, and therefore likely to be used mainly for global or summative analysis, omitting these topics is probably immaterial. However, if a form is lengthy, it would be hard to justify that gap.

Number of questions. The forms varied widely in terms of number of questions asked.⁶⁷ As Table 4 shows, current practice has not coalesced into any recognizable convention. [Table 4 is on next page.]

⁶⁷ Some forms contained questions applicable to all courses along with questions that apply only to particular kinds of courses, such as clinics. For this Article's analysis, only questions of general applicability to all courses were considered. In some instances, a law school's form contained questions mandated by a university's policies but apparently not intended to be used by the law school; those questions were ignored in this analysis.

Table 4: Number of Questionnaire Items
(percentages of schools)

Number of Items	Schools Using Forms within that range of number of items
5-9	18% ⁶⁸
10-14	29% ⁶⁹
15-19	21% ⁷⁰
20-24	16% ⁷¹
25-29	11% ⁷²
30 or more	5% ⁷³

⁶⁸ Barry, Campbell, Fordham, Idaho, Memphis, Nevada, Ohio State.

⁶⁹ Baltimore, Creighton, Georgia State, Houston, Loyola LA, Missouri (Kansas City), Missouri (Columbia), New England, UC Davis, Washburn, William Mitchell.

⁷⁰ Drake, Lewis and Clark, McGeorge, Regent, South Carolina, Suffolk, Thomas Cooley, University of Florida.

⁷¹ American, Catholic, Denver, Howard, Thomas Jefferson, Widener.

⁷² Nebraska, Southern Illinois, University of Utah, Vermont.

⁷³ CUNY, Stetson.

No particular number of questionnaire items can be described as typical. Apparently many schools are satisfied with the use of forms that contain fewer than ten items, while many schools choose to use forms that contain more than twenty items. Schools that use large numbers of items should be aware that data from global questions are likely to be more accurate than the sum of data from many more specific questions.

As the number of questions increases, so does the chance that those who use the data will be tempted to perform many different computations and comparisons, since they have so much more data to work with. This would contradict the main recommendation of this Article, that the data from student evaluation of teaching forms should be used to identify exceptional cases and not to make relatively small distinctions among instructors. The risk that using a large number of questions may create a false impression of greater accuracy might be worth running, if it allows a form to produce data on specific aspects of teaching that an institution considers particularly important. It seems, though, that sometimes there are not significant differences between forms with relatively low and relatively high numbers of questions. For an impressionistic consideration of this issue, Table 5 sets out each item from an eight-item form used by one school and each

item from a twenty-item form used by another school, attempting to match up the inquiries in each form. [Table 5 is on next page.]

Table 5: Comparison of Topics in Eight-Item and Twenty-Item Forms

Eight-item Form⁷⁴	Twenty-item Form⁷⁵
•Professor's knowledge of course material	•The instructor was familiar with materials assigned and appeared knowledgeable about the subject matter of the course. •The instructor had both practical and theoretical grasp of matters covered in the course. •The instructor demonstrated depth of knowledge of subject matter beyond the material assigned.
•Professor's organization	•The instructor was well-prepared and approached the course in an organized manner. •At the beginning of the course the instructor explained course objectives and informed students of the basis of grading and attendance policy.
•Professor's choice of class content	•Where applicable, the instructor related topics covered in the course to other areas of the law. •The instructor focused on the most important aspects of the course instead of relatively insignificant points. •The instructor adequately covered assigned material and major topics in the course.
•Professor's in class performance	•The instructor communicated ideas well, stimulated your interest in the subject matter, and challenged you to think critically about the course.
•Professor's relationship with students	•The instructor was accessible outside the class for additional discussion or guidance.
•Overall teaching ability	•The instructor was open to questions, handled them effectively, and focused discussions on topics relevant to the course. •The instructor made effective use of his or her chosen teaching methodology (lecture, Socratic dialogue, problems, etc.).
•Would you recommend that other students take this course?	•Your overall rating of this Instructor. •Your overall rating of this course.
•Additional comments	•Comments on instructor's knowledge of course: •Comments on instructor's teaching technique: •Comments on instructor's discharge of teaching duties: •General comments/suggestions:
	•Where applicable during the semester in which the course was offered, the instructor was reasonably prompt in posting grades, returning assignments, and supervision of written work or problems. (Ignore if not applicable).
	•The instructor rarely missed classes and met classes as scheduled.

⁷⁴ Fordham.

⁷⁵ Univ. of South Carolina.

In a number of instances, the longer form seems to include multiple and possibly redundant items on topics for which the shorter form uses just one item. For example, the shorter form asks about "professor's knowledge of course material," while the longer form covers "knowledgeable about the subject matter," "practical and theoretical grasp of matters covered," and "depth of knowledge of subject matter." Depending on the purpose for which the form is used and the ways in which users might analyze the data, the use of multiple items may lead to an impression of greater accuracy than could properly be justified.

Compound Questions. Individual items used in many forms use redundant or multiple expressions. Compound questions are a poor technique because answers to them may be ambiguous or because they prevent a clear answer when a respondent agrees with one term in the question but disagrees with another.⁷⁶

Examples of compound questionnaire items include:

"The professor was enthusiastic, stimulating, challenging, and thought-provoking."⁷⁷

"Presentation: clarity and organization."⁷⁸

"Demonstrating respect toward all students and fostering

⁷⁶ See, e.g., *Perez v. Z Frank Oldsmobile, Inc.*, 223 F.3d 617, 623-24 (7th Cir. 2000) (compound questions can make an answer ambiguous).

⁷⁷ American University.

⁷⁸ University of California, Davis.

mutual respect.”⁷⁹

“The instructor communicated ideas well, stimulated your interest in the subject matter, and challenged you to think critically about the course.”⁸⁰

“How successful was the instructor in exploring basic theoretical concepts and issues of public policy underlying this area of the law?”⁸¹

Open-ended and scaled questions. The forms use two types of questions: scaled and open-ended. A scaled question identifies an attribute and asks the student to evaluate that attribute by selecting one choice from a scale that might be expressed as ranging from strongly agree to strongly disagree or from excellent to unsatisfactory (typically in four or five point range). An open-ended question calls for a narrative response in the student’s own words.

The forms vary widely in whether they use open-ended questions, in their proportions of scaled and open-ended questions, and in whether their open-ended questions are general or guided. Overall, the forms included 534 scaled items and 144 open-ended items. Two forms used only open-ended questions.⁸²

⁷⁹ CUNY.

⁸⁰ University of South Carolina.

⁸¹ Creighton University.

⁸² Campbell University, University of Idaho.

One form used only scaled questions.⁸³ Forms with one or two open-ended questions were the most common. They were used at twelve schools.⁸⁴

Open-ended questions varied in the degree to which they guided the topics for responses. For example, one form⁸⁵ used the following three items:

“What has the instructor done especially well?”

“What should the instructor do to improve?”

“Comments: Please add comments on the above questions or on such matter as the instructions [sic] materials, improvement of the course, or any other subject:”

Using some open-ended questions increases the likelihood that the student evaluation of teaching data will not be used in a mechanistic way. The temptation to compare instructors by looking at small differences in their average scores is likely to be tempered when those scores are seen in the context of narrative descriptions of strengths and weaknesses.

Objective or Subjective Questions. For considering how various audiences use student evaluation of teaching data, it may be helpful to characterize the questionnaire items according to their relatively objective or subjective nature. An example

⁸³ University of Houston.

⁸⁴ Barry University, CUNY, Creighton University, Fordham University, Loyola Los Angeles, Lewis and Clark, New England, Ohio State, Regent, Thomas Jefferson, Denver, Widener.

⁸⁵ University of California, Davis.

of an objective item would be "The professor started and ended the class on time." An example of a subjective item would be "The professor was effective in allocating and using class time." Table 6 shows that many schools ask no objective questions. [Table 6 is on next page.]

Table 6: Number of Objective Questions
(percentages of schools using each number of questions)

Number of Objective Questions	Schools Using that many Objective Questions
0	63%
1	13%
2	8%
3	5%
4	3%
5	0%
6	0%
7	5%
8	3%

In the context of controversy about student evaluations, it is somewhat surprising that many schools fail to use the process to accomplish what would likely be the most readily accepted function, the collection of observations that do not involve judgment but might provide worthwhile information about basic aspects of teaching such as being punctual, providing a syllabus, or offering clear statements of students' obligations.⁸⁶

Explicit Assumptions. All of the forms include items like "effectiveness of teaching," or "presentation: clarity and organization." These items call for a general or summative conclusion about the student's experience. Their use represents institutional conclusions about courses and the work of professors - namely, that teaching should be effective or clear. Obviously, there is a consensus that teaching in all courses should have those attributes.

In contrast to that general agreement, some questionnaire items inquiries likely reflect individual schools' distinctive characteristics. For example, Regent Law School's form includes an item "Integrated Biblical principles into teaching the

⁸⁶ See Paul T. Wangerin, Evaluation of Teaching in Law Schools, 11 J. of Prof. Legal Education 87, 113 (1993) (suggesting that schools should "avoid asking questions that students cannot actually answer, questions, for example, regarding the degree to which a teacher is up-to-date...")

subject matter." Howard Law School includes "Did the course identify and discuss legal, social, economic and/or political problems that are of particular concern to marginalized groups?" in its form.⁸⁷

While all schools query about teaching in general and some schools use items that may be related to distinctive aspects of their missions, it may be noteworthy that attention to ethics and professionalism is relatively rare in the evaluation forms analyzed in this study. The recent Carnegie Corporation study on legal education concludes that in current law school teaching, ethical aspects of law receive too little attention.⁸⁸ When ethical considerations are incorporated in the curriculum, they are usually taught in a way that separates them from the concrete situations in which lawyers must face ethical problems and respond properly. Perhaps reflective of these Carnegie study conclusions is the fact that items referring to ethics appear in only five of the thirty-eight forms studied (Howard, Loyola Los Angeles, Lewis and Clark, Washburn, and William

⁸⁷ Other idiosyncratic questionnaire items are the following. American University's form includes "Instructor encouraged students to view the course material from different perspectives (e.g., different areas of the law and/or disciplines outside the law)." Vermont Law School's form includes "The professor effectively & sensitively addressed and responded to diversity issues." Loyola Los Angeles Law School's form includes "Making you aware of the human consequences of methods used or results sought by a lawyer."

⁸⁸ Sullivan, *Educating Lawyers: Preparation for the Profession of Law*, 31 (2007).

Mitchell). For example, the William Mitchell form asks "To what extent has this course prepared you to deal ethically with difficult situations in law practice," and the Washburn form asks students to "Comment on the professor's use or creation of opportunities in the classroom to discuss ethical problems relevant to the area of instruction."

Overall Review of Current Practices. The differences described above in number, style, and topic of questionnaire items and in the content of implied messages about education show that viewed across law schools, current practice manifests a wide range of approaches. The wide range of approaches is apparent even with regard to an attribute as simple as total number of inquiry items used. As shown above, the number of items used varies among schools so much that it would be impossible to state that any range of number of questions is typical. Considering styles of questions, open-ended questions are rare in some forms, a small proportion in many forms, and the only type of question used in another case. Some forms fail to avoid compound questions, despite the problems of interpretation that this format entails.

Topics in evaluation forms also vary. Most, but not all, ignore aspects of teaching that may well be significant to all users of the evaluation data, such as breadth of coverage and feedback. Topics on forms include some that are objective (such

as whether a syllabus was provided) and some that are subjective (such as "clarity") but there is wide variation in schools' decisions about how what proportion of the evaluation process should involve each category of inquiry.

Questionnaire items can convey fairly detailed messages about a school's view of the process of education. A minority of items focus students' attention on their own participation in their learning; most items implicitly treat the teaching-learning process as one in which the professor provides information or skills to passive recipients. Some forms delve specifically into the treatment of ethics in a course, but most do not, even though a consensus may exist that professional education should involve students in consideration of ethics much more than current practice does.

VI. Conclusion

Scholars who work in the field of education typically believe that student evaluations of teaching provide useful data that ought to be collected and considered for a range of purposes. In contrast, many law professors are skeptical about their value. This Article suggests that the discrepancy between these two views can be resolved, in part, by careful attention to the context in which student evaluation of teach data are used.

For administrators' decisions about faculty teaching assignments and students' course selection, even if student evaluation data has weaknesses, relatively imperfect information may be of use and does not carry any major risks of harm if it fails to be accurate. For improving teaching, current practices fail on two counts: the forms are vague and infrequent, while useful feedback for instructors would be specific and frequent. These shortcomings, fortunately, can be corrected by individual instructors on an ad hoc basis, by voluntarily using additional evaluation methods. For tenure and promotion decisions, schools should respect faculty perceptions that evaluations have weaknesses (and that some of those weaknesses are the result of students' biases). They should also respect the findings of education professionals that student evaluations are typically valid and reliable. These two goals can be harmonized. For tenure and promotion decisions, schools should use student evaluation of teaching data to identify outlier cases, where even if the data are affected by extraneous factors, the overall pattern of evaluations is quite far outside the institution's norms. Schools should be particularly aware that precise numerical comparisons between instructors or between different courses taught by a single instructor may often be statistically flawed.

Reviewing the thirty-eight forms that are the basis for

this Article's empirical analysis of forms shows that almost every questionnaire item conveyed or was consistent with the idea that learning is top-down, with professors providing and students receiving knowledge. The Article suggests that schools should protect students from this outmoded message.

Specific variations in the forms' topics and styles are plentiful. Some forms use many items and some use relatively few. Both approaches are supported in the education literature, but research has shown that global questions are better. It may also be true that detailed masses of questions can lead the users of the data to carry out precise arithmetic calculations and comparisons that may not be statistically legitimate. Use of open-ended questions also varies among schools. Open-ended questions can be valuable because they give specific reasons for respondents' numerical ratings and because including them may be a small defense against unthoughtful averaging or other mathematical comparisons.

Appendix: Schools Represented

School	US News Rank ⁸⁹		
	"Top 100" (rank)	"Tier 3"	"Tier 4"
American University	47		
Barry University			x
Campbell University			x
Catholic University	97		
City University of New York			x
Creighton University		x	
Drake Law School		x	
Fordham University	25		
Georgia State University	82		
Howard University		x	
Loyola Law School Los Angeles	66		
Lewis and Clark Northwestern	82		
McGeorge School of Law	100		
New England School of Law			x
The Ohio State University	31		
Regent University			x
Southern Illinois University		x	
Stetson College of Law	100		
Suffolk University		x	
Thomas Jefferson School of Law			x
Thomas M. Cooley School of Law			x
University of Baltimore			x
University of California, Davis	34		
University of Denver	77		
University of Florida	47		
University of Houston	60		
University of Idaho		x	

⁸⁹ <http://grad-schools.usnews.rankingsandreviews.com> (last visited February 8, 2007).

University of Memphis		x	
University of Missouri - Columbia	66		
University of Missouri - Kansas City		x	
University of Nebraska	77		
University of Nevada	100		
University of South Carolina	91		
University of Utah	57		
Vermont Law School		x	
Washburn University		x	
Widener University			x
William Mitchell College of Law		x	