University of Chicago

2015

The Teaching/ Research Trade-Off in Law: Data From the Right Tail

Tom Ginsburg, *University of Chicago*
Thomas J Miles

Article

The Teaching/ Research Trade-Off in Law: Data From the Right Tail

Ⓢ SAGE

Tom Ginsburg[1] and Thomas J. Miles[1]

Abstract

Background: There is a long scholarly debate on the trade-off between research and teaching in various fields, but relatively little study of the phenomenon in law. This analysis examines the relationship between the two core academic activities at one particular school, the University of Chicago Law School, which is considered one of the most productive in legal academia. **Method:** We measure of scholarly productivity with the total number of publications by each professor for each year, and we approximate performance in teaching with course loads and average scores in student evaluations for each course. In OLS regressions, we estimate scholarly output as a function of teaching loads, faculty characteristics, and other controls. We also estimate teaching evaluation scores as a function of scholarly productivity, fixed effects for years and course subject, and faculty characteristics. **Result:** Net of other factors, we find that, under some specifications, research and teaching are positively correlated. In particular, we find that students' perceptions of teaching quality rises, but at a decreasing rate, with the total amount of scholarship. We also find that certain personal characteristics correlate with productivity. **Conclusion:** The

[1] University of Chicago, Chicago, IL, USA

Corresponding Author:
Tom Ginsburg, 1111 E 60th, Chicago, IL 60637, USA.
Email: tginsburg@uchicago.edu

recent debate on the mission of American law schools has hinged on the assumption that a trade-off exists between teaching and research, and this article's analysis, although limited in various ways, casts some doubt on that assumption.

Introduction

The relationship between research and teaching is a topic of perennial discussion in academia. One camp argues that the two are substitutes, so that scholars will tend for the most part to be better at one or the other task (Wachtel 1998, 199). Another camp argues that the two are complements and that being a good teacher requires being a good scholar and vice versa (Wachtel 1998, 199). A third conceptual possibility is one of asymmetric complementarity, such that being a good teacher requires being a good scholar, but the inverse does not hold. Finally, it is possible that the two activities have no relationship at all.

While there is a substantial empirical literature on this question (see, e.g., Hattie and Marsh 1996 for a meta-analysis of 58 different studies), there are relatively few studies analyzing this trade-off in law (but see Merritt 1998; Barton 2008). The question is particularly important in light of a perceived gap between the activity of law schools and the actual training of lawyers (Edwards 1992; Katyal 2006). For decades, some have argued that law schools are becoming too divorced from legal practice and that they need to place less emphasis on faculty scholarship and more on skill-based training. This debate has become even more heated as law school tuitions have risen and the market for graduates have fallen (Tamanaha 2012). Some law deans have called for a fundamental restructuring of legal education (Rodriguez and Estreicher 2013), while others have vigorously defended the existing model. Recently, the American Bar Association's Task Force on the Future of Legal Education described the training of lawyers as currently riven by a "fundamental tension" between public and private values. The Task Force's leading example of this tension was the role of faculty scholarship in law schools. Faculty scholarship increases the cost of legal education, but the Task Force noted, "some argue that scholarship makes faculty members better teachers" (American Bar Association 2014, 7).

One of the main assertions has been that law schools have spent too many resources subsidizing research at the expense of training. Law students, it is said, are paying more in tuition partly to support higher faculty salaries and to fund more research (Campos 2011). Higher faculty salaries, in turn, have been dictated by the race to get the professors who are the best researchers, and to have a low student to faculty ratio in order to rise in the U.S. News annual rankings. Critics argue that, even if this model works for top tier schools, it does not necessarily work for others, but they follow the strategy of the top schools. All schools thus contribute to the ranking-related pathologies in the market.

The relationship between teaching and research in law schools is surely relevant for this heated debate. If there is a negative correlation between research and teaching in law schools, then it suggests that legal education should indeed focus less on research, as many of the critics claim. On the other hand, if there is a positive correlation, we can be more confident in the current model, at least in terms of relative allocation of resources between the two activities. Of course, even if there is a positive correlation, we cannot be sure that law schools are teaching the right things that students need to practice. That is a separate inquiry. The only issue we try to address here is whether there is a trade-off between the two core activities of the legal professorate.

The inquiry faces some significant measurement challenges. How does one quantify good teaching? As described subsequently in the section on data, we follow the literature in utilizing student ratings of professors as the indicator of teaching quality, though we are not confident that they accurately measure *quality* as opposed to popularity or some other dimension. Ideally one would have indicators of long-term absorption of subject matter as an indicator of teaching success, but such measures do not exist. On the other hand, teaching ratings are widely utilized and hence provide some advantages in terms of data availability and comparability. In particular, for our inquiry, teaching ratings allow for comparability within a single faculty, though they may be subject to specific biases related to discipline, department, or teaching institution that make comparison across faculties more difficult.

Measuring research quality is also subject to daunting problems. Our focus is on research productivity and so we focus on the quantity of scholarly output, as measured in a number of articles and books produced. We might think that citations, downloads, or national awards would be better indicators of the quality of a scholar. However, these measures are also biased in that they all favor more established scholars. We choose to focus

on productivity, though we acknowledge that it is quite possible that average article quality is lower for more productive scholars than it is for less productive scholars.

Our approach to measurement follows earlier studies. The most comprehensive recent study of the research–teaching nexus in law schools is Barton's (2008) study of 623 professors at 19 schools, primarily state institutions at a diverse range of levels in the academy. Utilizing publication and teaching data for every tenured or tenure-track faculty member at 19 American law schools for a 4-year period, the study argues that there is ''either no correlation or a slight positive correlation'' between teaching effectiveness and research productivity.

Barton's result is consistent with an earlier study by Merritt (1998) of 832 law professors. Merritt's sample was the entire population of law teachers who were between 7 and 11 years into teaching, including clinical and legal writing professors. Merritt's data were based on a survey with a response rate of over 50%. She examined as dependent variables total publications, publications in the top 20 law reviews, and books. Her indicators of teaching quality were teaching awards and total credit hours taught in any given year. She found a marginal, nonsignificant positive association between the number of articles published and winning a teaching award but no relationship between publishing in a top journal and an award. As she summarizes, ''This study provides no evidence of a relationship—either positive or negative—between teaching excellence and scholarly productivity'' (Merritt 1998, 808).

Merritt's study, valuable as it is, does not systematically control for the level of school. Barton's sampling techniques seem to be based on data availability and are not theoretically driven. We focus on a particular segment of legal academia: the ''right tail'' by which we mean the elite schools at the top of the hierarchy. We use our own law school, currently tied for fourth in the country according to the influential U.S. News ranking, as a case study, using more refined data than was available in the Merritt study. We do not assert that our school is representative of *all* legal academia. Indeed, even among top law schools, it may be exceptional on a number of dimensions, including productivity (Leiter 2000, 461–8). But this makes it a ''hard case'': If there is a negative relationship between research and teaching even at a top research school like Chicago, then it may call into question the very model to which other schools are striving. Furthermore, it is appropriate to examine the relationship between teaching and research at an elite institution because it is elite institutions that have placed the greatest emphasis on scholarship and on moving legal scholarship away from doctrinal and practice-oriented analyses.

The article is organized as follows: The next section examines the general literature on research and teaching in the academy. We then turn to our test in the Analysis section. Net of other factors, we find a weak relationship between teaching scores and scholarly output. In particular, we find that students' perceptions of teaching quality rise, but at a decreasing rate, with the total amount of scholarship. We also find that certain personal characteristics correlate with productivity. The analysis section concludes with the implications of the analysis for the recent debate on American legal education.

Research and Teaching: The Possible Relationships

Research and teaching are the two core activities of academic scholars, and different institutions take different approaches to encouraging one or the other. Much ink has been spilled asking whether one or the other is to be prioritized, and whether there is a trade-off between the two. Each camp has its adherents. At a conceptual level, there are four logical possibilities to describe the relationship between teaching and research: The two activities could have a negative relationship, a positive relationship, an asymmetric relationship, and no relationship. We examine each possibility in turn.

Negative Relationship

There are two main ideas that suggest a negative relationship between teaching and research. One idea is rooted in scarcity. This view draws on the scarcity of time and energy available to academics and the need to devote the marginal unit of effort to one or the other of the core activities. A second idea is that people might naturally be better suited to one or the other activity and that there is no reason to think that both skills will inhere in the same individual. In other words, the joint probability of having both skills is less than the probability of possessing only one of the skills on its own. Both scarcity and division of labor were well articulated by Newman as early as 1853, " . . . to discover and to teach are distinct functions; they are also distinct gifts, and are not commonly found united in the same person. He, too, who spends his day in dispensing his existing knowledge to all comers is unlikely to have either leisure or energy to acquire new" (Newman 1853, 10). Newman argued that the greatest thinkers spent little time on teaching but were instead devoted to the discovery of knowledge. From the point of view of an institution, both scarcity and specialization suggest that the teaching and research are competing in the production function: A unit

of input into one will diminish the other (Fox 1992). Some evidence for this can also be found in attitudes of researchers; a 1992 study showed that the most productive researchers are the least favorable toward teaching (Ramsden and Moses 1992, 273–95, quoted in Hattie and Marsh 1996, 510).

Positive Relationship

The contrary view is that the two core activities of academics are complementary. Hattie and Marsh (1996, 512) report that this view is taken by many academics themselves. It is commonly argued that research is necessary for an academic to keep up with his or her field (Hattie and Marsh 1996, 512). A researcher at the cutting edge, in turn, will be able to introduce her students to the latest ideas, ensuring that they are better prepared to apply them in practice. In this way, research may be directly complementary to teaching.

Asymmetric Relationship

Another logical possibility is that the two activities have different effects on each other. The intuition here is that, for example, a good teacher may or may not be a good researcher, but a good researcher is likely to be a good teacher. If good researchers tend to be more organized people than the average academic, a unit of time spent on research will not hurt their teaching. Alternatively, if teaching quality is determined in part by personal characteristics (such as empathy) that are uncorrelated with research, the two activities might have a weak relationship. Some evidence for the asymmetric position might be found in studies that show that time spent on teaching and research have different effects on outputs. Time spent on research is positively correlated with research productivity, while time devoted to teaching has little relationship with teaching quality (Hattie and Marsh 1996, 509 and 528).

No Relationship

Simply because time spent on teaching and research do not have equivalent returns does not mean the relationship between the two activities is asymmetric. Interpreting the data on returns to time invested in the two activities, Hattie and Marsh (1996, 509) argue that the actual correlation between teaching and research should be zero. The intuition here is that the time devoted to research, which takes away from teaching time, should not hurt

teaching performance because time does not contribute to that performance. A zero correlation would also follow if the two activities draw on different personality traits. Any observed positive relationship would be determined by the co-occurrence of the personal attributes rather than a causal effect of one activity on the other.

While there is no consensus, a large meta-study of 58 studies showed a small positive correlation between teaching and research (Hattie and Marsh 1996, 509). There was a negative effect between the year of publication of the primary studies and the teaching–research relationship, suggesting that even the small observed positive effect may be declining over time. The study found no difference across types of institution but did find that the field of study mattered: The positive relationship was greatest in the social sciences, followed by the humanities and the natural sciences (Hattie and Marsh 1996, 527). Law faculties tended to exhibit a slight negative relationship (Hattie and Marsh 1996, 527). Finally, the meta-study found a negative correlation between time spent on teaching and time on research. The scholars speculated that mediating variables, such as organizational skill, intellectual involvement in the teaching enterprise, and so on, might affect the relationship between teaching and research. Informed with these results, the same authors sought to examine mediating variables and found no relationship between teaching and research in a study of 187 faculty members (Hattie and Marsh 2004).

In short, the literature generally finds no relationship or only a very modest relationship between research productivity and teaching effectiveness, though some studies have found a weak positive relationship (Hattie and Marsh 2004, 199). The relationship seems to depend on discipline: A moderate positive effect has been found in social sciences but not the hard sciences (Centra 1983). But individual departments and universities differ. Furthermore, time and organization of academic activity may make a difference. For example, one public university forestry department analyzed cross elasticities and found that state-imposed increases in teaching obligations led to a decline in research productivity (Cubbage 1998).

Data and Method

Our data come from the University of Chicago Law School. The Law School is a relatively small one in terms of student enrollment and faculty size. Each cohort of juris doctor (JD) students typically consists of slightly fewer than 200 students, and the full-time tenure-track faculty, at this writing, consists of 34 members, a figure slightly higher than in most of the

years covered in this study. We analyze the data in two ways. First, we analyze the annual productivity of the faculty using an unbalanced panel of faculty members over the years 1999–2011. The years 2001, 2005, and 2006 are missing from the data due to an administrator's error. This loss of files leaves 10 years of data from a 13-year period. We do not, however, have any reason to think these missing years were unrepresentative. This panel is unbalanced because significant turnover marked the composition of the faculty during this period. Individual professors entered the data (through hiring) or exited the data (through departure for another law school, retirement, or death). The mean faculty member appears in the data for 5.1 years and the median for 5 years. Only three faculty members appear in each of the 10 years we observe, and six appear in 2 or fewer years. Overall, we observe 50 different faculty members during these years.

Second, we examine the average scores received in student course evaluations. For this analysis, we again study the years 1999–2011 with the caveat of the 3 missing years. But now the analysis occurs at the level of a course rather than a professor year. Here, the N is 496 courses, and again, we observe most of the professors repeatedly. The average professor taught an average of 10 courses and the median professor taught 9.

The curriculum in a 3-year JD program has a relatively standard structure across law schools. The first year consists almost entirely of mandatory courses, usually civil procedure and criminal law, along with common-law subjects such as contracts, property, and torts. The typical teaching approach uses the Socratic method, with the professor "cold calling" on a student to explain and justify (or critique) individual judicial decisions. During the second and third years, students choose their own courses that are either general demand electives or seminars.[1] Typically, general demand electives introduce students to substantive areas of law, such as "Public International Law," "Securities Regulation," or "Family Law," and instruction proceeds through a combination of lecture and Socratic method. By contrast, seminars have more narrowly focused topics, such as "Taxation of Derivatives," "European Legal History: From the 12th Century to Codification," or "Cicero's De Officiis (On Duties)," and classroom discussion is the primary pedagogy. At Chicago, enrollments in first-year courses are larger than in the average elective, commonly 90–95 students. Enrollments in seminars are capped at 25 students but are usually below the cap. The number of students in general demand electives varies widely from a dozen to more than 100. Of the 496 courses in our data, about 10% were mandatory first-year courses and 78% were general demand electives.

Teaching

We utilize student ratings of professors as the indicator of teaching quality, which is a standard, though imperfect, measure (Wachtel 1998; see also Rojstaczer 2012; McCulloch-Lovell 2012). A large debate in the academy pits critics of evaluations against supporters. Critics cite evidence that demonstrates bias on the basis of a set of nonverbal cues and qualities that seem unrelated to teaching (Merritt 2008, 238). As Merritt puts it, "the way in which a professor walks into the room or smiles at the class can affect student ratings much more substantially than what the professor says or writes on the blackboard" (Merritt 2008, 239). Supporters respond that there is no better way to evaluate teachers across a large range of departments and institutions and that teacher ratings do crudely capture student assessment (Barton 2008; Marsh 2007; Bok 2003). One survey summarizes several decades of research as "the majority of researchers believe that student ratings are a valid, reliable, and worthwhile means of evaluating teaching" (Wachtel 1998, 182; see Aleamoni 1999, 153–66; Heckert et al. 2006; Wright 2006; Marsh and Roche 1997).

While we have some sympathy with the critiques of teacher ratings and are not asserting that they are an indicator of "true" teaching effectiveness, we believe that teacher ratings have some advantage for our study. Most importantly, they allow for comparison across teachers and across time. We note that using ratings within a single institution, as we do, minimizes problems of comparability that arise in studies that aggregate multiple disciplines and faculties. Further, any alternative measure, such as teaching awards, is subject to their own difficulties that may make them worse measures of teaching performance. Teaching awards are rare events (at least at Chicago) and may introduce possible biases that make it difficult to compare more established with younger teachers.

The Law School has a standard protocol for soliciting teaching evaluations, one that is representative of the general approach in legal academia. Students are given a survey during the last week of class and asked to fill it out during class time. Instructors leave the classroom during the evaluation (Wachtel 1998, 195, explaining that ratings are somewhat higher when instructors are absent). Response rates vary but generally exceed 75% for large classes. The questionnaire consists of standard questions on the quality of the course, materials, and instructor. Students are asked to respond on a 5-point scale; earlier versions have used a 7- or 9-point scale. We normalized these scores to account for the differences across years. This might introduce some bias if students were inclined to view, for example, a

7-point scale in a different light than a 5-point scale.[2] For each class taught by a faculty member during the period under study, we calculated a weighted quality rating based on the average quality rating multiplied by the number of student responses. In the subsequent analyses, we deal with any potential bias resulting from this decision by disaggregating types of classes, distinguishing the large first-year classes from others.

This literature suggests certain control variables for our study. Wachtel (1998, 196–97) notes that class size matters: Smaller classes receives higher ratings, though some believe the relationship is curvilinear. There is also evidence that course characteristics might make a difference (Wachtel 1998, 198). The general literature suggests that instructors of higher rank (assistant, associate, full, or chaired professor) receive higher ratings, though age and experience do not have an independent effect (Wachtel 1998, 198). Indeed, some studies suggest that age and experience inversely correlate with ratings (Wachtel 1998, 198). Nor are there consistent gender effects (Wachtel 1998, 200; see generally, Laube et al. 2007; Sprague and Massoni 2005).[3]

Research

Our primary metric of research productivity comes from the annual Law School magazine for alumni that lists the scholarly publications appearing each year by faculty member. These include articles in law reviews and peer-reviewed journals, edited collections, and authored books. We distinguished between articles that appeared in a top 5 law review (defined by the most recent U.S. News ranking to include Yale, Harvard, Stanford, Columbia, and Chicago), a top 15 law review (as defined by U.S. News rankings of the #6 through #15 law schools), and other journals, including law reviews outside the top 15 and peer-reviewed journals.[4] To understand the relationship between teaching and scholarship, we create a single measure of scholarly output by summing the number of publications in each of these categories. We realize that this measure may not correspond to the level of time and effort required to produce each publication. For example, authoring a book may require more energy and time than a single article, and some articles surely require more efforts than others. But how much more is not clear. Moreover, books in legal academia often contain material that has previously been published as articles. As any weighting scheme would be arbitrary, we chose the simplest method and summed the total number of publications in all categories. This total is the primary variable

we analyze, but we assess its sensitivity by presenting results for individual categories of publication, as well.

We also present results in which articles in the popular press appear as a distinct category of publication. We do not treat these as part of a faculty member's scholarly output. Popular press articles are arguably a separate genre of writing with its own style, venues of publication, and largely non-academic readership. We exclude online writings, such as blog posts and "tweets," from our analysis altogether.

Covariates

Table 1 presents the data on teaching loads, scholarly output, and personal characteristics. During this period, about a quarter of the faculty were female and the median age was 49. More than 60% of the faculty were graduates of either Harvard or Yale law schools, which have produced the largest number of legal academics in the United States. A quarter of the faculty received PhDs from academic disciplines other than law. Before entering legal academia, more than half the faculty held at least one clerkship and more than half worked in legal practice. About 70% of the faculty were classified as "senior" in that they had held tenured positions at the time they appear in our data. About a third of them bore a heavy teaching load, and around 14% had heavy administrative responsibilities, which we define as serving as a dean, deputy dean, or chair of the faculty appointment committee. Chicago has long had a norm of placing few administrative duties on its faculty, with the exception of the appointments committee, in order that the faculty may devote more time to teaching and scholarship, and this is one of the ways in which our sample may be atypical. The high rate of turnover on the faculty is evident in that nearly 45% were lured to Chicago from another law school's faculty, and about a fifth departed for other law schools during our observation period.

Table 1 also shows that the faculty produces a substantial amount of scholarship annually. The average faculty member has nearly five scholarly publications per year. This total consists of about a third of a book, nearly a chapter in a collected volume, and three articles in law reviews. Of the law review articles, the average faculty member placed almost one of them in a top 5 law review.[5] As impressive as these averages are, there is substantial variation in the faculty's scholarly output. Some faculty had zero values in these categories in some years, and there is a long right tail in the distribution of output, which implies that a few faculty members produced very large amounts of scholarship. This is clear from the fact that the mean

Table 1. Summary Statistics: Faculty Teaching Loads, Scholarly Output, and Personal Characteristics.

Measure of Teaching or Scholarship	Mean	SD	Personal Characteristic of Faculty Member	Mean	SD
This year, number of …					
First-year courses taught	0.211	0.446	Female	.247	0.432
Electives taught (not seminars)	1.442	0.894	Age	47.41	11.243
Seminars taught	0.231	0.450	Graduate of Harvard Law School	.303	0.460
Student evaluations received	56.072	44.771	Graduate of Yale Law School	.323	0.468
Significant administrative duties this year	0.139	0.537	Had practice experience	.546	0.499
Articles in law reviews/journals …			Held Clerkship	.578	0.495
Outside top 15 or peer-reviewed	2.004	2.553	Received PhD	.255	0.437
In top 6–15	0.438	0.829	Junior Faculty Member	.315	0.465
In top 5	0.773	1.039	Hired laterally to Chicago	.438	0.497
Chapters in collected volumes	0.805	1.452	Subsequently left Chicago for another law school	.215	0.412
Collected volumes edited	0.135	0.407			
Books authored	0.311	0.764			
Articles in popular press	0.661	2.128			
Total scholarly output	4.474	4.815			

Note. SD = standard deviation.

number of total publications per year per faculty member is 4.63, while the median is 3. Another illustration is that one faculty member published 17 law review articles in a single year.[6]

The teaching scores faculty receive have two intriguing characteristics. The first is that students are asked to assign faculty a score between 1 and 5, and the mean score assigned is 4.31. This suggests that students think all of their faculty are excellent teachers. This may be so, given the emphasis that

Table 2. Summary Statistics: Student Evaluations of Teaching.

| Course Type	N	Normalized Score		Number of Student Responses	
		Mean	SD	Mean	SD
All	496	4.317	.485	28.375	24.355
First-year courses	53	4.213	.583	70.887	16.164
Electives (not seminars)	385	4.291	.474	25.584	20.177
Seminars	58	4.588	.359	8.052	4.110

Note. SD = standard deviation.

Chicago places on the quality of teaching, both as a matter of prevailing internal norms and with respect to the standards applied in hiring and promoting faculty. Alternatively, students assign scores to each faculty member individually and are not required to fit faculty teaching evaluations to a bell curve. For this reason, teaching evaluations may suffer from a form of "grade inflation."

The second intriguing aspect of the teaching scores is that they have a relatively small variance. The minimum score received is 3.07, and the maximum of 4.78, implying a range of only 1.71 relative to a potential scale of 1 to 5. This lack of variance may indicate that the scores students assign do not reflect careful judgments, or it may indicate the success of the law school's administration in securing high-quality teaching in most courses. Consistent with common intuitions, scores rise when students have choice over whether to enroll in a course, and they fall with course size. as a course's enrollment rises.

Table 2 also shows that students' perceptions of teaching vary with the type of course. Student ratings in mandatory first-year courses are the lowest and have the highest variance. Seminars receive the highest average evaluations and have the smallest variance. These patterns are consistent with a student preference for choice over mandates. The small enrollments in seminars enable more voluntary participation (as opposed to the involuntary "cold calls" of the Socratic method), and the narrower focus of seminars may facilitate a better matching of course subjects to student interests.

The number of student evaluations received also mirrors enrollment patterns. First-year courses receive the highest number of evaluations, but general demand electives have the highest variance. Seminars receive the fewest responses and the smallest variance. The patterns provide a hint that students are less likely to complete course evaluations during their second and third years.
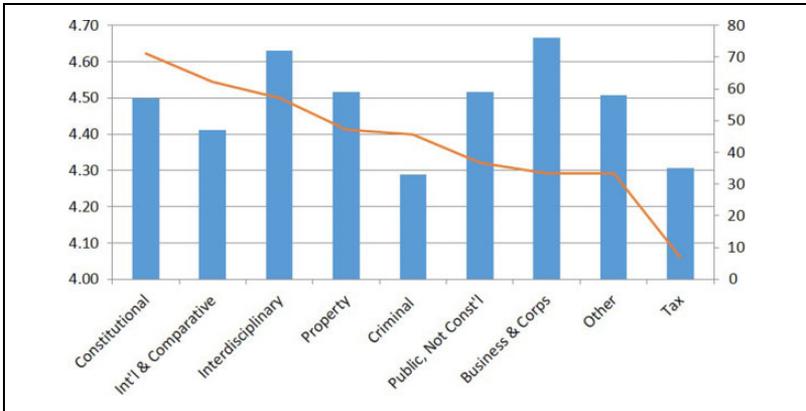
Figure 1. Student evaluations of teaching by course subject (median scores denoted by line on left scale, number of courses in sample denoted by bars on right scale).

Subject matter makes some difference for teaching ratings. Figure 1 shows the number of courses evaluated (right scale) and median scores received in those courses (left scale) by the subject area of the course. Constitutional law and international and public law courses receive the most favorable scores. By contrast, business law, corporate law, and tax law receive the lowest scores. A common intuition is that courses with complex statutes and transactions where questions are narrow in the sense that answers can be objectively correct (or wrong) are less enjoyable for law students than courses that deal with broad issues about the proper organization of society and in which answers depend more heavily on the quality of lawyerly argumentation. But we cannot, of course, reject the possibility that the latter type of course draws better teachers.

Analysis

Teaching Loads and Scholarly Productivity

Our primary concern is the relationship of teaching and research. As described in the data section, it is commonly thought that a trade-off between teaching and scholarship exists, and thus, a negative correlation should exist between scholarly output and the scores received in teaching evaluations. The data and method section also advanced reasons why

teaching and scholarship might be positively correlated or not correlated at all. Raw correlations suggest a very weak relationship: the two are correlated at only .02, but numerous factors may influence this relationship. Therefore, regression analyses are appropriate.

First, we estimate the relationship between the amount of teaching (or teaching loads) and scholarly productivity. Then, we turn to the relationship between scholarly output and students' perceptions of teaching quality. To control for the numerous factors potentially influencing the amount of scholarship, we estimate equations of the form

$$Y_{it} = X_{it}\beta + W_{it}\delta + \alpha_t + \varepsilon_{it},$$

where Y_{it} is a count of professor i's scholarly output in year t. The term X_{it} contains measures of each professor's teaching and administrative responsibilities in each year, and the term contains W_{it} a set of personal characteristics of each faculty member. The term α_t is a fixed effect for year t. Finally, ε_{it} is an error term, and standard errors are clustered by faculty member.

The dependent variable, Y_{it} or the number of scholarly publications a professor produced in a given year, is a form of count data, which suggests that a Poisson regression is the preferred estimation technique. We report here results from ordinary least squares regressions because the coefficients lend themselves more readily to interpretation than Poisson. In unreported regressions, results from Poisson estimation were similar.

For an initial cut at the data, we simply sum all forms of scholarly output into a single measure of scholarly productivity. As described above, given that the relative effort required to produce books and articles is not clear, we arbitrarily assign equal weights to articles, chapters in collected volumes, books edited, and books authored in our summary measure. After examining the relationships of this measure to teaching, we consider different forms of scholarship individually.

Table 3 examines whether higher teaching loads reduce scholarly productivity. Teasing out the impact of teaching responsibilities on scholarship is challenging because the assignment of faculty to courses is not random. For instance, faculty may request to teach particular courses, and where possible, these requests are honored. At other times, curricular needs compel the administration to request faculty to cover a course that that the faculty member does not typically teach. Also, the administration seeks to maintain a high quality of teaching in required first-year courses and will shift faculty in and out of these courses when their performance warrants it.

Table 3. OLS Estimates: Overall Scholarly Output as a Function of Teaching Loads and Personal Characteristics.

Explanatory Variable	Total Scholarly Output				Popular Press	
	(1)	(2)	(3)	(4)	(5)	(6)
This year, number of …						
First-year courses taught	1.433 (1.682)	.477 (1.495)	−.255 (1.715)	−1.243 (1.853)	.808 (1.009)	.510 (.643)
Electives taught	−.070 (.421)	.113 (.549)	−.131 (.512)	.050 (.738)	.235 (.204)	.318 (.209)
Seminars taught	.243 (.894)	.544 (.733)	−.333 (.784)	−.018 (.758)	.236 (.422)	.353 (.388)
Student evaluations received	−.0048 (.0148)	.0087 (.0142)	.0100 (.0148)	.0263 (.0197)	−.0007 (.0075)	.0020 (.0057)
Significant administrative duties this year	.300 (.406)	−.136 (.433)	.492 (.382)	.134 (.447)	−.243 (.165)	−.399 (.253)
Last year, number of …						
First-year courses taught	—	—	.979 (1.525)	.086 (1.564)	—	—
Electives taught	—	—	−.189 (.886)	−.101 (1.022)	—	—
Seminars taught	—	—	−.778 (.877)	−.261 (.804)	—	—
Student evaluations received	—	—	−.0085 (.0117)	.0093 (.0175)	—	—
Significant administrative duties last year	—	—	−2.157* (1.171)	−1.758* (.9116)	—	—
Female	—	−1.345 (2.002)	—	−.371 (2.756)	—	−.994 (.721)
Age	—	.170 (.315)	—	.028 (.414)	—	−.044 (.111)
Age squared	—	−.0007 (.0034)	—	.0005 (.0043)	—	.0008 (.0011)
Graduate of Harvard Law	—	9.38 (1.274)	—	.442 (1.524)	—	.395 (.444)

(continued)

61

Table 3. (continued)

Graduate of Yale Law	—	.444 (1.646)	—	-.259 (2.124)	—	.890 (.770)
Had practice experience	—	.099 (1.362)	—	.063 (1.628)	—	-.303 (.630)
Held clerkship	—	.237 (.982)	—	-.252 (1.317)	—	.346 (.390)
Received PhD	—	.615 (1.872)	—	.874 (2.065)	—	1.0957 (.845)
Junior faculty member	—	-.351 (1.285)	—	-.764 (1.293)	—	-.180 (.674)
Hired laterally to Chicago	—	2.226 (1.694)	—	2.978 (1.996)	—	1.112 (.829)
Subsequently left Chicago for another law school	—	1.970 (1.300)	—	2.266 (1.386)	—	.495 (.518)
Adjusted R^2	.0083	.0791	.0460	.0335	.0131	.0957
N	251	251	178	178	251	251

Note. OLS = ordinary least squares. Robust standard errors in parentheses.
***$p < .01$, **$p < .05$, *$p < .1$

62

The regression in column (1) of Table 3 estimates scholarly productivity as a function of teaching loads in the current year (and year fixed effects). The regression in column (2) adds measures of the personal characteristics of the faculty. These two regressions do not show a clear relationship between teaching loads and scholarly productivity. A striking feature of Table 3 is that almost none of the estimates are statistically significant. Most of the coefficients are positive rather than negative, and the signs of two variables reverse when the faculty's personal characteristics are included. The estimates do not support the existence of a trade-off between scholarship and teaching loads.

The magnitudes of most of the coefficients imply modest impacts on the count of scholarly output. The coefficient implying the largest effect is that in column (1) for the number of first-year courses taught. Taken at face value, it suggests that faculty who teach one first-year course produce more than one additional scholarly item each year, but the inclusion of personal characteristics in column (2) cuts this estimate by two thirds. This suggests that the identity of the individual faculty member, rather than the course-load, is the stronger predictor of scholarly output. Also, it implies that the administration's assignment of more productive faculty to first-year courses may dominate any productivity reduction from the responsibility of teaching first-year courses with their relatively large enrollments and greater student demands for faculty attention.

Columns (3) and (4) include 1-year lags of teaching and administrative responsibilities. There is usually some delay between the production of scholarship and its final publication. As a rough attempt at capturing this delay, we add measures of the prior year's responsibilities to the regressions. Most of the coefficients on these variables are negative, and their inclusion drives the coefficients on two measures of the current year's teaching duties from positive to negative. Particularly noteworthy are the statistically significant negative coefficients for administrative duties. Those who serve the institution in an administrative capacity pay a price in terms of productivity.[7]

For the most part, we postpone discussion of the relationship of a faculty member's personal characteristics and scholarly output until the next tables, where we examine particular categories of scholarship. But it is worth noting here the personal characteristics with the largest coefficients. Those are whether the faculty member was hired laterally to Chicago and whether the faculty member was hired laterally from Chicago (i.e., departed to teach at another law school). Both of these experiences correlate with a higher level of scholarly productivity. They imply that faculty members who are active

in the lateral market produce on average about two more pieces of scholarship annually. This is consistent with the conventional understanding that faculty with the most lateral opportunities tend to be those who produce more scholarship. (Note that we have no data on the productivity of faculty before or after they teach at Chicago.)

The final two columns of Table 3 provide a comparison of scholarship and writing for the popular press, which we excluded from our category of scholarly output. One view is that writing for the popular press is not an academic activity, that it is a distraction from the core academic enterprise of research and teaching. An alternative view is that it benefits an academic institution to have some faculty who are public intellectuals because their visibility draws favorable attention to the institution and to the work of all its faculty. Much legal scholarship has direct policy implications, and participation in the public conversation through popular press writing may help instantiate academic ideas into policy.

The regressions in columns (5) and (6) have annual counts of a professor's popular press articles as their dependent variables. These estimates are largely similar to those on scholarly output in columns (1) and (2). The coefficients are smaller in magnitude because the average number of popular press articles per year is 0.661, compared to 4.474 for scholarly items. But most of the coefficients bear the same sign as in the prior regressions, and the relative sizes of the coefficients within each regression are similar. These patterns confirm that production of scholarship and popular press writing is correlated. Several explanations may exist for this pattern. One possibility is that prolific writers may produce a greater volume of all forms of writing, both academic and popular. Another possibility pertains less to the nature of the writing skill but the incentives for writing for the public. Some faculty, especially at Chicago, may place the highest priority on academic writing. They may then see writing for the general public as of drawing attention to their scholarly work, and faculty who produce more scholarship will have more occasions to write for the general public.

In Table 4, we decompose the aggregate scholarship measures and consider the relationship between individual categories of scholarship and teaching loads and faculty characteristics. The categories of scholarship are articles in journals and law reviews ranked below 15, law reviews ranked 6 through 15, law reviews ranked 5 and above, chapters in collected volumes, books edited, and books authored. Examination of the specific types of scholarly outputs does not provide a clearer pattern of the relationships between scholarly productivity and teaching loads. For example, the

Table 4. OLS Estimates: Specific Categories of Scholarly Output as a Function of Teaching Loads and Personal Characteristics.

Explanatory Variable	Journals and Law Reviews Outside Top 15		Top 6–15 Law Reviews		Top 5 Law Reviews	
	(1)	(2)	(3)	(4)	(5)	(6)
This year, number of ...						
First-year courses taught	.235 (.950)	−.352 (.763)	.256 (.201)	.301 (.193)	.137 (.333)	.255 (.311)
Electives taught	−.123 (.278)	−.101 (.341)	.026 (.088)	.073 (.091)	−.014 (.091)	.068 (.094)
Seminars taught	.309 (.464)	.451 (.380)	−.092 (.148)	−.126 (.138)	−.123 (.131)	−.119 (.123)
Student evaluations received	−.0013 (.0080)	.0069 (.0073)	−.0011 (.0019)	−.0016 (.0017)	.0027 (.0025)	.0007 (.0030)
Significant administrative duties this year	−.097 (.313)	−.338 (.380)	.142* (.086)	.140 (.091)	.162 (.130)	.087 (.145)
Female	—	−.743 (1.028)	—	−.381** (.156)	—	−.666*** (.244)
Age	—	.011 (.155)	—	.036 (.060)	—	.055 (.057)
Age squared	—	.0002 (.0016)	—	−.0003 (.0006)	—	−.0005 (.0006)
Graduate of Harvard Law	—	.099 (.617)	—	.245 (.235)	—	.362 (.265)
Graduate of Yale Law	—	.340 (.944)	—	.138 (.163)	—	.174 (.221)
Had practice experience	—	−.154 (.723)	—	.041 (.166)	—	.073 (.192)
Held clerkship	—	.069 (.575)	—	−.039 (.124)	—	.076 (.163)
Received PhD	—	−.130 (1.039)	—	−.243 (.182)	—	−.063 (.287)
Junior faculty member	—	−.0096 (.6684)	—	−.015 (.253)	—	.250 (.334)
Hired laterally to Chicago	—	1.497* (.893)	—	.130 (.227)	—	.160 (.317)
Subsequently left Chicago for another law school	—	.521 (.573)	—	.595* (.328)	—	.350 (.303)
R^2	.0501	.1506	.0360	.1602	.0981	.2140

(continued)

Table 4. (continued)

Explanatory Variable	Chapters in Collected Volumes		Collected Volumes Edited		Books Authored	
	(1)	(2)	(3)	(4)	(5)	(6)
This year, number of . . .						
First-year courses taught	.579* (.354)	.058 (.412)	.030 (.151)	.015 (.116)	.173 (.225)	.174 (.179)
Electives taught	−.029 (.142)	−.060 (.182)	−.016 (.048)	.012 (.048)	.080 (.060)	.118* (.060)
Seminars taught	.097 (.236)	.157 (.213)	−.031 (.077)	.010 (.068)	.073 (.155)	.155 (.120)
Student evaluations received	−.0059 (.0032)	.0025 (.0040)	.0013 (.0017)	.0010 (.0014)	−.0004 (.0029)	−.0007 (.0021)
Significant administrative duties this year	.172 (.185)	.120 (.115)	−.035 (.045)	−.072 (.046)	−.039 (.055)	−.066 (.057)
Female	—	.506 (.541)	—	−.074 (.057)	—	.008 (.204)
Age	—	.022 (.087)	—	.0062 (.025)	—	.041 (.045)
Age squared	—	.0000 (.0001)	—	.0000 (.0003)	—	−.0002 (.0005)
Graduate of Harvard Law	—	−.023 (.360)	—	.177 (.092)	—	.095 (.163)
Graduate of Yale Law	—	−.317 (.395)	—	.075 (.061)	—	.035 (.154)
Had practice experience	—	−.064 (.338)	—	.048 (.064)	—	.129 (.147)
Held clerkship	—	−.046 (.307)	—	.074 (.070)	—	.111 (.129)
Received PhD	—	.798* (.468)	—	.129* (.079)	—	.102 (.166)
Junior faculty member	—	−.249 (.320)	—	−.057 (.085)	—	−.273* (.152)
Hired Laterally to Chicago	—	.528 (.372)	—	−.016 (.069)	—	−.069 (.156)
Subsequently left Chicago for another law school	—	.379 (.250)	—	−.050 (.053)	—	.186 (.136)
R^2	.0664	.2558	.0635	.1660	.0493	.1715

Note. OLS = ordinary least squares. Robust standard errors in parentheses.
*$p < 0.1$, **$p < .05$, ***$p < .01$.

66

number of seminars taught appears positively related to articles in law reviews ranked below 15 but negatively related to articles in law reviews ranked 15 and above. This is puzzling because the effort cost of producing a law review article likely does not depend on whether it is placed in a top law review, and the differences in quality between articles in law reviews ranked just above and just below 15 are likely small.

In addition, some estimates are highly sensitive to whether the regression controls for the characteristics of faculty members. The relationship of the number of first-year courses taught to articles in law reviews below rank 15 swings from positive to negative once professorial characteristics are introduced. Similarly, its relationship to chapters in collected volumes falls by an order of magnitude.

The correlations between faculty characteristics and scholarly productivity vary widely across the type of scholarship. Female faculty members were less likely to publish law review articles but more likely to publish chapters in collected volumes. Production of all categories of scholarship rose with age, but the increase was linear for some categories (e.g., chapters in collected volumes) and concave for others (e.g., articles in highly ranked law reviews). The point estimates imply that graduates of Harvard Law School were more likely to publish in highly ranked law reviews, while graduates of Yale Law School were more likely to publish in lower ranked law reviews.

As interdisciplinary scholarship has grown, a common concern is that interdisciplinary research will be misguided if academics lack practical knowledge of how legal institutions function. In faculty hiring decisions, this sometimes leads to discussions as to whether an interdisciplinary candidate is ''law enough,'' and the types of legal experience most commonly valued are a period of legal practice and having held an appellate clerkship. The results in Tables 4 show that faculty with practice or clerkship experience appear no different from other faculty in the volume of their scholarly productivity. The magnitudes of their point estimates are generally small relative to those of other variables, and some estimates even bear negative signs, implying that scholars with a closer connection to legal practice produce less scholarship.

But there is some evidence that possession of a PhD, the standard marker of extensive training in another discipline, alters where a faculty member chooses to publish. Professors with a PhD appeared less likely to publish in law reviews but more likely to contribute to collected volumes. (We do not breakout peer-reviewed journals in our analysis.) This pattern is consistent with a common impression about PhD holders in legal academia:

that they prefer to publish in venues that are edited by other academics and are less enamored of publishing in student-edited law reviews.

The results in Tables 4 provide more detail on the higher productivity of scholars who recently changed employers. Faculty who departed Chicago appear to have produced more law review articles, especially in top reviews. By contrast, faculty who joined Chicago as lateral hires were more productive in publishing articles outside of the very top law reviews. Perhaps the most surprising feature of the results is that junior faculty (those who have not yet received tenure) do not appear to produce a different total volume of scholarship than senior faculty.

At many academic institutions, one piece of conventional wisdom is that scholarly productivity falls once a professor receives tenure. The conventional explanation relies on incentives: That professors work less hard when the risk of job loss is removed. With the exception of one category of scholarship, this prediction is not borne out in the law school at Chicago. The summary statistics on scholarly output in Table 1 provide some support for Chicago's reputation of having a highly productive faculty. Perhaps the selection of faculty or the culture of the school overcomes the incentive effects of tenure. The sole category of scholarship in which the junior faculty were less productive was the authoring of books. In legal academia, a book is often a compilation and reworking of previously published articles. In order to recapitulate prior work in book form, a scholar must have prior work to draw upon, and junior faculty at the beginnings of their careers are less likely to have this stock of prior creations.

Students' Perceptions of Teaching and Scholarly Productivity

So far, our examination of the potential trade-off between teaching and scholarship has focused on whether a larger quantity of teaching responsibility comes at the price of a lower volume of scholarship. It is also possible that professors sacrifice the quality of teaching in order to increase scholarly output. In this section, we test this hypothesis by analyzing the relationship between teaching quality and the volume of scholarship. Specifically, we estimate teaching quality as a function of scholarly productivity and other characteristics using regressions of the form:

$$S_{jiqt} = g(Y_{it}) + X_{it}\rho + W_{it}\theta + Z_j\eta + \lambda_t + w_{jiqt},$$

where S_{jicqt} is the average score on student evaluations professor i received in course j in quarter q and year t. The term $g(Y_{it})$ is a function of a professor i's scholarly output in year t. We report results from various

functional forms of $g(Y_{it})$, such as linear and quadratic. As before, the term X_{it} contains measures of each professor's teaching and administrative responsibilities in each year, and the term W_{it} contains a set of personal characteristics of each faculty member. The term Z_j contains measures of the course characteristics, such as indicators for the subject areas shown in Figure 1. It also contains the number of student evaluations received and the nature of the instruction (e.g., a required first-year course or a seminar). The terms λ_q and λ_t are fixed effects for quarter q and year t, respectively. Finally, ω_{it} is an error term, and standard errors are clustered by faculty member.[8]

Table 5 reports the regression estimates. The regressions in the first three columns include just measures of scholarly productivity, course characteristics, and teaching loads. The regressions in the last three columns add controls for professor characteristics. The simplest measure of scholarship, a count of the number of publications, suggests a weak relationship to the quality of teaching. In columns (1) and (2), the estimated coefficients are less than .02. But when scholarly output enters the equation as a quadratic, a more complex relationship emerges. In columns (2) and (4), student perceptions of teaching have a positive but concave relationship with scholarly output. Moreover, the positive relationship dominates for the reasonable and perhaps feasible range of scholarly output. The point estimates imply that the level of output that maximizes student perceptions of teaching is extraordinarily high, about 16 publications per year (!). This level of output is not attainable for all but the most exceedingly prolific scholars and, thus, the estimates suggest that for most law professors, the quality of teaching rises with the total volume of scholarship. The remaining columns decompose the total count of scholarly output into the counts of its component categories. Here, the largely positive relationship between scholarly productivity and teaching quality remains. In column (3), five of the six categories of scholarly output have positive estimates, and in column (6), four of six do.

In addition, the positive relationship between teaching quality and scholarly quantity appears important in terms of its magnitude. The estimates for scholarly productivity in columns (2) and (4) are not only statistically significant, they imply that a rise in scholarship translates into a meaningful increase in student perceptions of teaching. For example, an increase in scholarly output of 1 standard deviation (*SD*; or roughly 4.8 publications) corresponds to an increase of about one quarter of a *SD* in teaching scores according to the estimates in column (2), and slightly less than one half of a *SD* according to the estimates in column (4).

Table 5. OLS Estimates: Student Evaluations of Teaching as a Function of Teaching Loads, Scholarly Output, and Personal Characteristics.

Explanatory Variable	(1)	(2)	(3)	(4)	(5)	(6)
Total scholarly output	.012 (.008)	.032* (.019)	—	.016* (.009)	.044*** (.016)	—
Total scholarly output squared	—	−.0010 (.0007)	—	—	−.0013** (.0006)	—
Articles in law reviews . . .						
Outside top 15	—	—	−.022 (.014)	—	—	−.011 (.011)
In top 15	—	—	.050* (.029)	—	—	.041 (.036)
In top 5	—	—	.019 (.030)	—	—	−.002 (.028)
Chapters in collected volumes	—	—	.024 (.018)	—	—	.048** (.019)
Collected volumes edited	—	—	.086** (.042)	—	—	.075 (.055)
Books authored	—	—	.075** (.036)	—	—	.051 (.035)
Articles in popular press	−.015 (.011)	−.016 (.010)	−.012 (.012)	−.009 (.008)	−.011 (.007)	−.005 (.008)
Number of student evaluations received	−.0031** (.0013)	−.0032** (.0013)	−.0029** (.0014)	−.0045*** (.0012)	−.0047*** (.0012)	−.0044*** (.0012)
Multiple courses taught this quarter	−.106 (.067)	−.102 (.067)	−.096 (.067)	−.093 (.064)	−.090 (.064)	−.087 (.066)
Number of courses taught this year	.010 (.062)	.001 (.064)	−.017 (.056)	−.022 (.051)	−.035 (.054)	−.039 (.052)
Significant administrative duties this year	.068* (.040)	.057 (.041)	.065* (.034)	.065* (.035)	.056 (.038)	.059** (.028)

(continued)

Table 5. (continued)

First-year course	.073 (.151)	.079 (.151)	.049 (.149)	.081 (.131)	.092 (.133)	.068 (.128)
Seminar	.193*** (.063)	.199*** (.064)	.203*** (.064)	.121* (.064)	.125* (.064)	.123* (.064)
Female	—	—	—	.083 (.067)	.114* (.070)	.050 (.066)
Age	—	—	—	.050** (.026)	.048* (.025)	.048** (.024)
Age squared	—	—	—	−.0006** (.0002)	−.0005** (.0002)	−.0006** (.0002)
Graduate of Harvard Law	—	—	—	.160* (.064)	.150** (.064)	.151** (.059)
Graduate of Yale Law	—	—	—	.046 (.075)	.044 (.073)	.050 (.069)
Had practice experience	—	—	—	−.047 (.053)	−.058 (.055)	−.051 (.051)
Held clerkship	—	—	—	.120* (.073)	.126* (.070)	.113 (.073)
Received PhD	—	—	—	.053 (.082)	.033 (.081)	.020 (.078)
Junior faculty member	—	—	—	−.241** (.087)	−.250*** (.089)	−.221** (.084)
Hired laterally to Chicago	—	—	—	−.284** (.074)	−.299*** (.069)	−.267*** (.074)
Subsequently left Chicago for another law school	—	—	—	−.011 (.076)	−.005 (.075)	−.024 (.077)
R^2	.1690	.1749	.1911	.2801	.2903	.2931

Note. OLS = ordinary least squares. Robust standard errors in parentheses.
*$p < .1$, **$p < .05$, ***$p < .01$.

71

These estimates are surprising in view of the conventional wisdom that there is a trade-off between teaching and scholarship. Even if teaching and scholarship were complements, one might expect that time constraints would compel a trade-off between teaching and research. The other interesting finding is that types of scholarship differ. It is interesting that publications in law reviews do not have a stronger relationship to teaching. In contrast, the point estimates for books and chapters are larger. These results imply that a professor who writes most often for student-edited journals on average does not receive the highest evaluations from those students. Instead, faculty who write for venues more favored by other academics perform more strongly in the class room.

The positive relationship between total scholarly output and students' perceptions of teaching quality has several potential explanations. One is that teaching and scholarship are actually complements. But as just mentioned, it is surprising that the type of scholarship which is most likely to complement legal pedagogy (i.e., law review articles) does not have a positive linear relationship to student perceptions of teaching. Another explanation may be that students favor faculty who are more prominent academics, and faculty who publish more garner higher profiles. Yet, this account runs into difficulty when one contrasts the estimates for scholarly output and articles in popular press. Professors who write most often in the popular press may enjoy a type of fame, but the estimate for popular press articles is negative and small in absolute value. (Although not shown in the table, when the regression includes a quadratic of popular press articles, the estimated coefficients on both of its terms are negative and smaller in absolute value than .01.) A third possibility, which we are unable to test with these data, is that Chicago students prefer academically inclined faculty and these are faculty who publish more peer-reviewed articles.

The results in Table 5 also suggest that the nature of the course influences students' perceptions of teaching quality. Upper year elective courses (the omitted category) are less popular than seminars but approximately as popular as mandatory first-year courses. Several features of seminars may account for their higher evaluations. They have smaller enrollments (capped at 25), they typically proceed with open discussion rather than cold calling, they allow more interaction with the faculty, and their readings are often varied and encompass materials other than the standard legal texts of statutes, regulations, and judicial opinions. Furthermore, students self-select into these courses, whereas many other ''elective'' courses are considered to be advisable for preparation for the bar exam. The scores given by students also fall as the number of student evaluations per course

increases. This may be due to the fact that students are more likely to complain about poor teaching than to compliment excellent teaching. But the patterns in Table 2 suggest that the size of a course's enrollment accounts for most of the variation in the number of evaluations received. Thus, the appropriate interpretation may be that the quality of teaching suffers as enrollment rises. But even here, the impact is modest. The estimates imply an additional 10 student evaluations received imply a reduction of less than .05 of point, which is about a tenth of a *SD* in the scores.

Several of a faculty member's characteristics correlate with the scores received. Age has an increasing but concave relationship to the scores. This is consistent with increasing but diminishing returns to experience in teaching. The fact that junior faculty regularly receive lower scores is also consistent with the idea that experience matters in teaching. The point estimates imply that faculty in their 40s receive the highest scores. Administrators tend to be highly regarded as teachers. Although graduates of Harvard and Yale account for the lion's share of legal academics, the estimates suggest they do not perform appreciably better in the classroom. Faculty who received their law degrees at Harvard received slightly higher scores, about .15 of a point, and the correlation for attendance at Yale is about a third of this. The traditional forms of legal experience present conflicting correlations, with practice experience implying lower student evaluations and clerkships implying higher ones.

Faculty who depart Chicago for other law schools appear similar in their teaching evaluation scores to those who remain at Chicago. By contrast, professors who were hired laterally to Chicago (a category that constitutes about 40% of the faculty-years in the sample) after teaching at other schools receive lower teaching evaluations. This may reflect that scholarly productivity is more valued than teaching in lateral hiring or that teaching quality is less readily observed at the time of lateral hiring decisions. Another possibility is that professors who have been active in lateral markets place a higher priority in scholarship than teaching or that the teaching techniques developed at other law schools do not match well with Chicago's students.

As described earlier, the content and structure of law courses vary widely by their type. First-year courses are mandatory, have large enrollments and cold calling, and cover core legal doctrines. Seminars have small enrollments, student discussion, and focus on more specific, often interdisciplinary and academic, topics. General demand electives have variable enrollments, combine lecture and cold calling, and tend to focus on substantive law. The student experience in these types of courses is likely to differ, and thus one might expect that the relationship of scholarly output and

faculty characteristics to students' perceptions of teaching might differ by type of course. Table 6 explores this possibility by estimating regressions for teaching evaluations separately for each type of course.

Several differences appear by the type of course. The positive, concave relationship between scholarly output and teaching quality occurs in courses with upper year law students: electives and seminars. The relationship is much weaker in mandatory first-year courses. A potential explanation for this pattern is that students require a foundation of knowledge about law in order to appreciate the value that a faculty member's scholarship brings to their teaching, and first-year students lack this foundation.

The negative relationship between the number of reviews received and teaching quality is most persistent in elective courses. This pattern provides further support for interpreting this measure as a proxy for the number of students enrolled in a course, because enrollments in mandatory first-year courses vary very little and seminars by construction have low enrollments.

All of the regressions in Table 6 suggest that teaching multiple courses in a single term may reduce the quality of teaching. The estimate for seminars is particularly large (in absolute value), and this pattern may hint at the importance of time constraints. A common practice is for faculty members to teach seminars on subjects closely related to their current research, and, accordingly, they may teach new seminars more frequently than new courses. If a faculty member is preparing a new seminar while also teaching other courses, time constraints might impair the quality of the seminar.

Faculty demographics appear to matter. Most prominently, female faculty members receive lower average scores in first-year courses but higher scores in seminars. This pattern is consistent with some of the existing literature that shows gender effects in teaching evaluations (Wachtel 1998, 200; Laube et al. 2007; Sprague and Massoni 2005). But caution should be taken against attaching too much import to these estimates, as we have not sought to develop a complete analysis of gender in our sample. Faculty members who graduated from Harvard Law School receive slightly higher scores in first-year and elective courses, but they perform less well in seminars. Yale graduates on the faculty fare even more poorly in seminars.

Faculty with PhDs receive substantially lower scores in first-year courses. But it is not obvious that this pattern is an indictment of interdisciplinary scholars teaching doctrinally intense courses, because faculty with experience in legal practice also receive lower scores in those courses (although the magnitude of the estimate is smaller). The experience of a clerkship correlates most closely with higher teaching scores in first-year

Table 6. OLS Estimates: Student Evaluations of Teaching as a Function of Teaching Loads, Scholarly Output, and Personal Characteristics.

Explanatory Variable	First-Year Courses		Electives (not Seminars)		Seminars	
	(1)	(2)	(3)	(4)	(5)	(6)
Total scholarly output	.010 (.103)	—	.052*** (.017)	—	.076 (.049)	—
Total scholarly output squared	−.0014 (.0032)	—	−.0015** (.0007)	—	−.0019 (.0014)	—
Articles in law reviews …						
Outside top 15	—	−.040 (.044)	—	.0009 (.0139)	—	−.014 (.091)
In top 15	—	.065 (.197)	—	.042 (.032)	—	.316 (.417)
In top 5	—	−.322* (.172)	—	.044 (.025)	—	.037 (.241)
Chapters in collected volumes	—	.177* (.101)	—	.035 (.021)	—	.056 (.132)
Collected volumes edited	—	.346 (.342)	—	.094 (.068)	—	.283 (.234)
Books authored	—	.038 (.123)	—	.036 (.031)	—	.122 (.208)
Articles in popular press	−.046 (.066)	−.113** (.058)	.0002 (.010)	.004 (.008)	−.008 (.051)	−.037 (.071)
Number of student evaluations received	.006 (.008)	.006 (.009)	−.0050*** (.0014)	.0047*** (.0014)	−.0055 (.0324)	−.010 (.038)
Multiple courses taught this quarter	−.379 (.389)	−.128 (.392)	−.043 (.066)	−.041 (.064)	−1.252 (.829)	−1.743 (1.036)
Number of courses taught this year	.269 (.366)	.165 .255	−.113 (.059)	−.122** (.057)	.189 (.361)	.256 (.378)
Significant administrative duties this year	.109 (.169)	.184 (.243)	.104*** (.027)	.104*** (.029)	−.070 (.122)	−.014 (.091)
Female	−.571* (.311)	−1.067*** (.273)	.119* (.070)	.095 (.076)	.342 (.216)	.327 (.249)
Age	.128 (.091)	.082 (.090)	.041 (.028)	.042 (.027)	−.050 (.103)	−.034 (.118)
Age squared	−.0012 (.0008)	−.0007 (.0009)	−.0005* (.0003)	−.0005 (.0003)	.0007 (.0011)	.0004 (.0013)
Graduate of Harvard Law	.186 (.400)	.680* (.349)	.168* (.062)	.155** (.063)	−.147 (.369)	−.148 (.539)
Graduate of Yale Law	−.005 (.478)	.619 (.409)	.081 (.069)	.068 (.069)	−.515 (.383)	−.459 (.558)

(continued)

75

Table 6. (continued)

Had practice experience	−.397* (.236)	−.349 (.277)	−.074 (.062)	−.063 (.058)	−.062 (.296)	−.120 (.307)
Held clerkship	.459 (.353)	.187 (.271)	.080 (.081)	.092 (.085)	.410* (.212)	.499** (.183)
Received PhD	−.549 (.610)	−1.377** (1.539)	.009 (.083)	.020 (.080)	−.225 (.172)	−.285 (.331)
Junior faculty member	.276 (.385)	.690* (.355)	−.300*** (.096)	−.299*** (.090)	−.084 (.509)	−.116 (.576)
Hired laterally to Chicago	.114 (.407)	−.043 (.408)	−.284*** (.067)	−.254*** (.076)	−.447 (.494)	−.440 (.460)
Subsequently left Chicago for another law school	.410 (.511)	.482 (.427)	−.018 (.082)	−.026 (.087)	−.054 (.318)	−.016 (.330)
N	53	53	385	385	58	58
R^2	.6760	.7649	.3299	.3271	.5533	.6560

Note. OLS = ordinary least squares. Robust standard errors in parentheses.
*$p < .1$, **$p < .05$, ***$p < .01$.

76

courses and seminars. The lower average scores received by junior faculty, as reported in Table 5, are due primarily to their performance in upper year electives. By contrast, junior professors receive higher scores than their senior colleagues in first-year courses.

Discussion

There has been much debate about whether the performance of American law schools is failing, according to various metrics. Critics argue that U.S. law schools subsidize useless research, taking resources away from the core activity of educating students. Our own school is alleged to be a part of the problem, as it is part of a set of elite schools that "set the market" for the legal professorate. A key assumption of these critics is that a trade-off exists between teaching and scholarship: That the time and energy spent on research impair the quality of teaching or that professors proficient at scholarship (or scholarship that is not directly related to legal practice) are poor teachers of law.

The results of our study cast doubt on this assumption, at least at one elite school. Our analysis of 10 years of data encompassing nearly 50 faculty members over almost 500 courses reveals no strongly negative relationship between the volume of scholarship and the amount or quality of teaching. To the contrary, scholarly output appears to rise with the amount of teaching in many specifications, although the relationships are not statistically significant. In addition, students' evaluations of teaching have a positive and statistically significant correlation with the feasible range of total scholarly output. The estimates suggest that rather than restricting scholarly activity, teaching quality might be improved by reducing class enrollments and the administrative burdens on faculty.

The article's results also address the specific claim that faculty who engage in interdisciplinary scholarship are poor teachers of courses with a heavy content of legal doctrine. Faculty with the traditional indicia of legal experience—appellate clerkships and legal practice—do not receive consistently higher teaching evaluations. Nor do they publish more often in law reviews than do other faculty members. Faculty with PhDs in other disciplines receive on average less favorable teaching evaluations in first-year courses, but their teaching ratings are indistinguishable from their colleagues. Yet, they tend to publish more often, especially in collected volumes. These results provide very little support for the view that interdisciplinary scholars are not a good fit for the mission of law schools.

The results of the analysis are, of course, limited in important ways. The analysis draws on data from only one school, a school that is atypical in its ranking and in its academic culture, with a reputation for intense intellectual engagement. The analysis is correlational because the assignment of professors to courses is not random. There are additional and more precise measures of scholarship and teaching quality that one might prefer. For these reasons, we do not assert that our results undermine the critique for all schools. However, our results do suggest that, for most faculty members at one elite law school, there is a positive correlation between teaching and research prowess.

Acknowledgment

Declaration of Conflicting Interests

Funding

Notes

1. We do not analyze clinical courses in which second- and third-year students engage in apprentice-like forms of legal practice under the guidance of an experienced attorney who is usually a clinical professor. These courses differ from traditional academic teaching, and clinical faculty often do not publish scholarly articles.

2. There is some evidence in the literature that respondents to surveys view different scales differently. To test for this possibility, we examined the variance on the normalized scale over time and found that it did not change systematically depending on the nominal scale that was in place. We further note that all our reported regressions include fixed effects for years, which should absorb much of the annual variation in the dependent variables. Finally, in unreported analyses, we ran our primary regressions on subsets of the data for each different scale. We found no major substantive differences from our reported results, except that

the 10-point scale was in use for so few observations in our sample that no results proved to be significant.

3. But note ''there was a significant interaction between gender, attractiveness and authoritarianism; namely, teachers with an authoritarian philosophy were rated less negatively if they were attractive and female'' (Wachtel 1998, 200).

4. Definitions of "top" law reviews are inevitably arbitrary. We include Harvard, Yale, Stanford, Columbia, and Chicago in the top five. Our top 15 grouping includes these 5 plus New York University, the University of Pennsylvania, California, Virginia, Cornell, Georgetown, Michigan, Northwestern, Texas, and University of California, Los Angeles law reviews.

5. We note that the top 5 includes the University of Chicago Law Review. Some argue that law reviews tend to favor their own faculty.

6. Although not shown here, we assessed the sensitivity our analyses of the determinants of teaching and scholarship to these outlying values and concluded that they did not drive our conclusions.

7. The years of lost teaching evaluations imply that a sizable number of values of last year's teaching duties are missing. Accordingly, the sample size is smaller in columns (3) and (4), and for this reason, we should limit our attention to the current year's teaching duties.

8. We also ran models with fixed effects for professors and found no substantive differences in the results.

References

Aleamoni, L. M. 1999. "Student Rating Myths versus Research Facts from 1924 to 1998." *Journal of Personnel Evaluation in Education* 13:153–66.

American Bar Association Task Force on the Future of Legal Education. 2014. *Report and Recommendations*. January 24. Accessed February 2, 2014. http://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/report_and_recommendations_of_aba_task_force.authcheckdam.pdf.

Barton, Benjamin. 2008. "Is There a Correlation between Law Professor Publication Counts, Law Review Citation Counts, and Teaching Evaluations? An Empirical Study." *Journal of Empirical Legal Studies* 5:619–44.

Bok, Derek. 2003. *Universities in the Marketplace*. Princeton, NJ: Princeton University Press.

Campos, Paul. 2011. *The Cost of Legal Scholarship*. Accessed May 15, 2013. http://insidethelawschoolscam.blogspot.com/2011/11/cost-of-legal-scholarship.html.

Centra, J. A. 1983. "Research Productivity and Teaching Effectiveness." *Research in Higher Education* 18:379–89.

Cubbage, Frederick W. 1998. "An Empirical Example of Tradeoffs between Teaching and Research." *Natural Resources and Env'tl Issues* 7:210–11.

Edwards, Harry T. 1992. "The Growing Disjunction between Legal Education and the Legal Profession." *Michigan Law Review* 91:34–70.

Fox, Mary Frank. 1992. "Research, Teaching and Publication Productivity: Mutuality versus Competition in Education." *Sociology of Education* 65:293–305.

Hattie, John, and H. W. Marsh. 1996. "The Relationship between Research and Teaching: A Meta-analysis." *Review of Educational Research* 66:507–42.

Hattie, John, and H. W. Marsh. 2004. *One Journey to Unravel the Relationship between Research and Teaching.* Paper presented at Research and Teaching: Closing the Divide? An International Colloquium. http://www.education.auckland.ac.nz/webdav/site/education/shared/hattie/docs/relationship-between-research-and-teaching-(2004).pdf

Heckert, Teresa M., A. Latier, A. Ringwald-Burton, and C. Drazen. 2006. "Relations among Student Effort, Perceived Class Difficulty Appropriateness, and Student Evaluations of Teaching: Is It Possible to "Buy" Better Evaluations through Lenient Grading?" *College Student Journal* 40:588–96.

Katyal, Neal Kumar. 2006. "Hamdan v. Rumsfeld: The Legal Academy Goes to Practice." *Harvard Law Review* 120:65–123.

Laube, Heather, Kelley Massoni, Joey Sprague, and Abby L. Ferber. 2007. "The Impact of Gender on the Evaluation of Teaching: What We Know and What We Can Do." *National Women's Studies Association Journal* 19:87–104.

Leiter, Brian. 2000. "Measuring the Academic Distinction of Law Faculties." *Journal of Legal Studies* 29:451–93.

Marsh, H. W. 2007. "Students' Evaluations of University Teaching: A Multidimensional Perspective." In *The Scholarship of Teaching and Learning in Higher Education: An Evidence-based Perspective*, edited by R. P. Perry and J. C. Smart, 319–84. New York: Springer.

Marsh, H. W., and Lawrence A. Roche. 1997. "Making Student's Evaluations of Teaching Effectiveness Effective." *American Psychologist* 52:1187–97.

McCulloch-Lovell, Ellen. 2012. *As Long as Its about More Than Scores.* Accessed January 30, 2014. http://www.nytimes.com/roomfordebate/2012/09/17/professors-and-the-students-who-grade-them/student-evaluations-work-as-long-as-theyre-about-more-than-scores.

Merritt, Deborah Jones. 1998. "Research and Teaching on Law Faculties: An Empirical Exploration." *Chi.-Kent Law Review* 73:765–822.

Merritt, Deborah Jones. 2008. "Bias, the Brain, and Student Evaluation of Teaching." *St. John's Law Review* 82:235–87.

Newman, John Henry. 1853. Accessed January 30, 2014. *The Idea of a University Defined and Illustrated: In Nine Discourses Delivered to the Catholics of Dublin.* http://www.newmanreader.org/works/idea/preface.html

Ramsden, P., and I. Moses. 1992. "Association between Research and Teaching in Australian Higher Education." *Higher Education* 23:273–95.

Rodriguez, Daniel B., and Samuel Estreicher. 2013. "Make Law Schools Earn a Third Year." *NY Times*, January 13. Accessed January 30, 2014. http://www. nytimes.com/2013/01/18/opinion/practicing-law-should-not-mean-living-in-bankruptcy.html?_r=0.

Rojstaczer, Stuart. 2012. "Bad Data That Leads to the Wrong Answer." *NY Times Online Debate*. Accessed January 30, 2014. http://www.nytimes.com/roomfor-debate/2012/09/17/professors-and-the-students-who-grade-them/student-eva luations-offer-bad-data-that-leads-to-the-wrong-answer.

Sprague, Joey, and Kelley Massoni. 2005. "Student Evaluations and Gendered Expectations: What We Can't Count Can Hurt us." *Sex Roles: A Journal of Research* 53:779–93.

Tamanaha, Brian. 2012. *Failing Law Schools*. Chicago, IL: University of Chicago Press.

Wachtel, Howard K. 1998. "Student Evaluation of College Teaching Effectiveness: A Brief Review." *Assessment and Evaluation in Higher Education* 23:191–212.

Wright, Robert E. 2006. "Student Evaluations of Faculty: Concerns Raised in the Literature, and Possible Solutions." *College Student Journal* 40:417–22.

Author Biographies

Tom Ginsburg is the Leo Spitz Professor of International Law and Ludwig and Hilde Wolf Research Scholar at the University of Chicago Law School. He holds JD and PhD degrees from the University of California at Berkeley.

Thomas J. Miles is the Clifton R. Musser Professor of Law and Economics and Walter Mander Research Scholar at the University of Chicago Law School. He holds a JD degree from Harvard Law School and a PhD in economics from the University of Chicago.