

Is There a Correlation between Scholarly Productivity, Scholarly Influence and Teaching Effectiveness in American Law Schools? An Empirical Study

Benjamin Barton¹

This empirical study attempts to answer an age-old debate in legal academia: whether scholarly productivity helps or hurts teaching. The study is of an unprecedented size and scope. It covers every tenured or tenure-track faculty member at 19 American law schools, a total of 623 professors. The study gathers four years of teaching evaluation data (calendar years 2000-03) and creates an index for teaching effectiveness. This index was then correlated against five different measures of research productivity. The first three measure each professor's productivity for the years 2000-03. These productivity measures include a raw count of publications and two weighted counts. The scholarly productivity measure weights scholarly books and top-20 or peer reviewed law review articles above casebooks, treatises or other publications. By comparison, the practice-oriented productivity measure weights casebooks, treatises and practitioner articles at the top of the scale. There are also two measures of scholarly influence. One is a lifetime citation count, and the other is a count of citations per year.

These five measures of research productivity cover virtually any definition of research productivity. Combined with four years of teaching evaluation data the study provides a powerful measure of both sides of the teaching versus scholarship debate. The study correlates each of these five different research measures against the teaching evaluation index for all 623 professors, and each individual law school. The results are counter-intuitive: there is no correlation between teaching effectiveness and any of the five measures of research productivity. Given the breadth of the study, this finding is quite robust.

Anyone who has spent any time in legal academia has heard some version of the scholarship versus teaching debate. The debate breaks down into two camps that I call “pro-teaching” and “pro-scholarship” for brevity’s sake, although I recognize that either camp may object to these labels as too simplistic. The pro-teaching folks bemoan how current legal academia places excessive emphasis on scholarly pursuits, and argues that we are thus inevitably short-changing our students. The pro-scholarship group retorts that our best scholars are naturally our best and most up-to-date teachers. Thus, it is the

¹ Associate Professor of Law, University of Tennessee College of Law. B.A. 1991, Haverford College; J.D. 1996, University of Michigan. The author gives special thanks to Indya Kincannon, Glenn Reynolds, Eugene Volokh, James Lindgren, Brian Leiter, James Maule, Richard E. Redding, Brannon Denning, Tom Galligan, Joan Heminway, Mae Quinn, Greg Stein, Jennifer Hendricks, Geroge Kuney, Jeff Hirsch, Chris Sagers, the participants of faculty forums at the University of Texas Law School, Villanova University School of Law, Cumberland School of Law, Samford University, the University of Tennessee College of Law, the Southeastern Association of Law Schools Panel on Empirical Research, the University of Tennessee College of Law for generous research support, and the Honorable Diana Gribbon Motz.

faculty that neglects scholarship that is actually harming students.² This debate has also been echoed in various law review articles.³

If either the pro-scholarship or pro-teaching faction is right, there are serious implications for the design and management of legal academia. Many pro-teaching folks argue that the scholarship/teaching balance is seriously skewed in current law schools, to the detriment of students and the public at large. If their argument is correct, it suggests that law schools should shift their focus and energies towards more teaching-centric activities. If the pro-scholarship argument is correct, law schools should pour even more energy into their scholarly missions. Their faculty will be on the cutting edge of their fields, and the students will reap the benefits.

Interestingly, the debate does not primarily rest on philosophical or ideological grounds. To the contrary, both sides of the debate make what are essentially empirical claims. As always in this sort of debate, both sides cloak their empirical claims in very plausible stories.

On the one hand, it seems likely that working hard on scholarship should have some positive effect on teaching. Productive scholars do tend to stay on top of their research areas, and are also very engaged with the material they write about. On the other hand, it also makes sense that if law professors are spending more and more of their time on scholarship they must be shortchanging their teaching.

Which side is right? To answer this question I attempt to empirically measure whether there is a correlation between teaching effectiveness and either research productivity or scholarly influence (as measured by a citation study). The study actually reaches a conclusion unlikely to please either side of the debate: on five different measures of research productivity, and across the tenured and tenure-track faculty at 19 different American law schools there is no correlation between teaching evaluations and research productivity or influence.

The study is easily the largest and most thorough of its kind in American law schools or at any level of American higher education. In addition, the study's extraordinary access to four years of teaching evaluation data at 19 American law schools, as well as the usage

² My friend and colleague Brannon Denning has reduced this argument (and this study's response) to the following puppet show script:

Pig One: "Hooray! I am an exceptional scholar, and perforce a great teacher!"

Pig Two: "Sadly I fear you are mistaken. Time is a scarce resource. Since you spend all your time writing law review articles you are short-changing your students. Place a pox upon the US News and the scholarship arms race."

[Enter **Pig Three**, with charts] **Pig Three:** "I bring glad tidings! You are both wrong, the data shows no correlation."

[**Pigs One and Two** examine the charts] **Pigs One and Two** in unison: "Arrgh! Is it too late to go work for a law firm?"

³ See, e.g., Russell Korobkin, *In Praise of Law School Rankings: Solutions to Coordination and Collective Action Problems*, 77 TEX. L. REV. 403, 422-25 (1998) (pro-scholarship); Marin R. Scordato, *The Dualist Model of Legal Teaching and Scholarship*, 40 AM U. L. REV. 367, 368-83 (1990) (pro-teaching).

of five different measures of scholarly productivity makes it arguably the most comprehensive and authoritative study of its kind in any university setting.

I. PREVIOUS STUDIES

Over the last 50 years there have been a number of studies of the correlation between teaching effectiveness and research productivity in higher education. There are three excellent overviews of these studies: a meta-analysis by John Hattie and Herbert Marsh (1996), John Braxton's synthesis (1996), and Kenneth Feldman's earlier collection of studies (1987).⁴ Overall, these three overviews establish that there is either no correlation or a slight positive correlation between research productivity and teaching.

As a matter of methodology the various studies underlying Hattie and Marsh, Braxton, or Feldman are something of a mixed bag. Many rely upon self-reporting for either teaching effectiveness or scholarly productivity, and most focus upon a single institution or department. Also, since many are self-reported, they rely upon the portion of the total studied population that chose to respond to the study. The time frames studied are also typically much shorter than four years.⁵

Hattie and Marsh sought to avoid many of these problems in their more recent study of 182 Australian university professors. Their thorough study used multiple measures of research productivity, and also made a strong finding of no correlation.⁶ I followed many

⁴ In some regard each of these articles involve a "meta analysis" of the previous research/teaching studies in that they each take a broad and mathematical look at all the studies to reach an overall conclusion on the research/teaching correlation. See generally Jeremy A. Blumenthal, *Meta Analysis: A Primer for Legal Scholars* (draft on file with author) for a description of meta-analyses.

The first article to analyze the previous studies of teaching and research is John A. Centra, *Research Productivity and Teaching Effectiveness*, 18 RES. HIGHER ED. 379 (1983), which conducted an original study and compared it to previous studies and reported that based on the available data there was either no correlation, or the correlation was too modest to draw any firm conclusions. See *id.* at 385-88. Kenneth Feldman, *Research Productivity and Scholarly Accomplishment of College Teachers as Related to Their Instructional Effectiveness: A Review and Exploration*, 26 RES. HIGHER ED. 227 (1987) took a much more comprehensive look at the then existing studies and found either no correlation or a small positive correlation. See *id.* at 275-80 ("[I]n general, the likelihood that research productivity actually benefits teaching is extremely small or [the] two, for all practical purposes, are essentially unrelated.").

Two articles in the 90s reviewed the past studies, plus the studies completed after Feldman's analysis, and reached a firmer finding of no correlation. John Hattie and H.W. Marsh performed a meta analysis of 58 studies and found no correlation. See John Hattie & Herbert W. Marsh, *The Relationship Between Research and Teaching: A Meta-Analysis*, 66 REV. ED. RES. 507,529-30 (1996) ("We must conclude that the common belief that research and teaching are inextricably intertwined is an enduring myth."); see also John M. Braxton, *Contrasting Perspectives on the Relationship Between Teaching and Research*, in FACULTY TEACHING AND RESEARCH: IS THERE A CONFLICT? (John M. Braxton ed., 1996) (finding the "lack of a systematic relationship between teaching and research" after a review of 28 studies).

⁵ Herbert W. Marsh & John Hattie, *The Relation Between Research Productivity and Teaching Effectiveness: Complementary, Antagonistic, or Independent Constructs?*, 73 J. HIGHER ED. 603, 606-17 (2002) offers a good overview of some of these methodological shortcomings.

⁶ See Marsh & Hattie, *supra* note __, at 628-37 (finding a "clear" indication "that teaching effectiveness and research productivity are nearly uncorrelated").

of the methodological suggestions of Hattie and Marsh in this study. Notably, I studied every tenured or tenure-track faculty member at 19 different institutions over 4 calendar years, and none of the data are self-reported.

There are two prior law school studies of the correlation between research productivity and teaching effectiveness. Deborah Jones Merritt studied whether there was any correlation between law school teaching awards and scholarly productivity, and concluded that there was no statistically significant correlation.⁷ Merritt's excellent study is far-reaching, and remarkably broad (it includes 832 law professors) and covers many subjects outside of the teaching/scholarship correlation. Nevertheless, Merritt herself recognizes that her teaching award data involved self-reporting and might have been unreliable.⁸ Furthermore, using teaching awards as a proxy for teaching effectiveness is somewhat problematic, since these awards are handled differently at every institution.

James Lindgren and Allison Nagelberg conducted a second study.⁹ It looked at 48 professors (16 professors each from 3 law schools -- the University of Chicago, the University of Colorado, and Boston University). The 16 faculty members selected were the 8 most cited faculty members and the 8 least cited faculty members at each school from a separate citation study by Theodore Eisenberg and Martin T. Wells.¹⁰ Teaching evaluations were gathered for each of these faculty members, and Lindgren and Nagelberg did a correlation study. They found a statistically significant correlation of .20, and also that the most highly cited professors were more likely to have higher teaching evaluations than the least cited professors.¹¹

While the Lindgren and Nagelberg study is well done, it is on a smaller scale than this study, and tracks only three relatively non-diverse institutions. Further, the choice to eliminate the middle of the faculty in terms of citations makes the conclusions less concrete than a study that covers the entire faculties of 19 different law schools. Lindgren and Nagelberg chose to study the extremes on each faculty because of the uncertainty of citation studies as a measure of scholarly influence.¹² Nevertheless, if there is a concern over the accuracy of a single measure of scholarly influence, it is better to try multiple measures than to eliminate a substantial chunk of the studied faculties.

II. METHODOLOGY

In 2003 I set out to study whether there is a correlation between teaching effectiveness and scholarly productivity in American law schools. I planned to gather teaching evaluation data for a four year period (2000-2003) from 20 American law schools, and

⁷ See Deborah Jones Merritt, *Research and Teaching on Law Faculties: An Empirical Exploration*, 73 CHI.-KENT L. REV. 765, 781-83 (1998).

⁸ See *id.* at 772-74.

⁹ See James Lindgren & Allison Nagelberg, *Are Scholars Better Teachers?*, 73 CHI.-KENT L. REV. 823 (1998).

¹⁰ See Theodore Eisenberg & Martin T. Wells, *Ranking and Explaining the Scholarly Impact of Law Schools*, 27 J. LEG. STUD. 369 (2000).

¹¹ See Lindgren & Nagelberg, *supra* note __, at 827-32.

¹² See *id.* at 826-27.

then correlate this data with a study of faculty productivity over the same period of time. I wrote a prospectus describing the project, and mailed the prospectus and a cover letter to the deans of forty American law schools.¹³

The response rate was significantly below the 50% I had hoped for. I ended up writing every dean and associate dean of every ABA or AALS accredited law school in America, which only resulted in 13 schools agreeing to participate. I then turned to State Freedom of Information Act requests, and eventually gathered data from 19 total law schools. Because the process of gathering the teaching evaluations was so difficult, and reaching my goal of 20 law schools would have likely required a public information law suit, I decided to go forward with 19 rather than 20 law schools.

A. Who Was Studied

The nineteen law schools studied are:

- 1) The University of Colorado School of Law;
- 2) The University of Connecticut School of Law;
- 3) The Cumberland School of Law, Samford University;
- 4) The Levin College of Law at the University of Florida;
- 5) The University of Iowa College of Law;
- 6) The Lewis & Clark Law School;
- 7) The University of Michigan Law School;
- 8) The University of North Dakota Law School;
- 9) The Northwestern University School of Law;
- 10) The Moritz College of Law at the Ohio State University;
- 11) The Penn State Dickinson School of Law;
- 12) The Southwestern Law School;
- 13) St. John's University School of Law;
- 14) The University of Tennessee College of Law;
- 15) The Texas Tech University School of Law;
- 16) The University of Toledo Law School;
- 17) The UCLA School of Law;
- 18) The Villanova University School of Law;
- 19) The Wayne State University Law School.

These 19 law schools are an excellent and representative sample. It includes schools from every region of the country, every level of academic reputation, every size, and a balance of public and private institutions. Given that there are only 193 ABA accredited law schools¹⁴ gathering four years of data on every tenured and tenure-track faculty member from almost 10% of the total number of American law schools offers a comprehensive sample.

¹³ A copy of the original prospectus can be found at <http://www.law.utk.edu/FACULTY/BartonProp.PDF>.

¹⁴ ABA, Section of Legal Education and Admission to the Bar, ABA-Approved Law Schools, <http://www.abanet.org/legaled/approvedlawschools/approved.html> (last visited June 20, 2006). For the AALS list, see AALS, Member Schools, http://www.aals.org/about_memberschools.php (last visited June 20, 2006).

A total of 623 faculty members were studied from these 19 schools. I generated a list of the tenured or tenure-track faculty for each of those schools during the study period. I included only tenured and tenure-track faculty because at many schools there is not a requirement of, or support for, scholarly activities of the library, clinical, writing, or adjunct faculty. Because these faculty members are not required or encouraged to publish many do not. Including those faculty members thus could have skewed the analysis.¹⁵

There were, of course, faculty coming and going during a four-year period, so I included any faculty member who had teaching evaluations from at least four different classes during the time period. This ensured that each professor had a fair number of responses, and also allowed the study to cover as many faculty members as possible.

B. What Was Studied -- Teaching Effectiveness Data

The teaching evaluation data came in different formats for different institutions, from access to a university website that gathered the data, to a single-page amalgamation, to physical copies of every student evaluation during the period. From these data, I chose the question on the evaluation sheet that most closely measured teaching effectiveness. For example, The University of Tennessee College of Law's form actually asks the students to rank the professor from 1-5 (with 5 being the highest ranking) on the "Instructor's effectiveness in teaching material." The results can be found on a publicly accessible website (University of Tennessee 2006). Of the 19 schools, 13 schools asked a somewhat similar question and ranked the professor from 1-5. Two of the other schools ranked from 5-1 (with 1 being the best ranking), one ranked from 4-1 (again with 1 as the best), and one each ranked from 1-4, 1-7, and 1-9, with 1 being the lowest.

I then took the teaching evaluation data for each professor and averaged the data over the four-year period. If the number of students was given I weighted each evaluation by the number of students in the class, so that within any professor's results a seminar of 7 students was counted differently from a large class. The number of student responses was provided for 15 of the 19 schools, so most of the averages are weighted.

I weighted the data as follows. For each individual class I multiplied the average student ranking on teaching effectiveness by the number of student responses. I then added this total amount from each class and divided the number by the total number of students. That average number was the average evaluation *per student*. I could have just added the averages per class, and then divided by the number of classes to generate the average evaluation *per class*. Averaging the classes without adjusting for class size would mean the average ranking of students in a small seminar or independent study would weigh the same as the ranking of 75 students in a lecture class, greatly discounting the input of the

¹⁵ Nevertheless, a non-scientific review of these teaching evaluations suggests that their inclusion would not have greatly changed the results of the study.

students in the lecture class. Because I thought that teaching effectiveness should be reflective of the overall effect on students I decided to adjust by class size.¹⁶

I then converted each professor's average teaching evaluation score into a teaching evaluation index that ranged from 0-100. I did this so that I could make an overall correlation comparison between teaching and research productivity or influence for all of the professors studied. The index was calculated as follows. For the 13 schools that ranked professors from 1-5 I used the following equation:

$$\frac{\text{Teaching evaluation} - 1}{4} \times 100$$

By definition this score will range from 0-100. I subtracted one from both the numerator and denominator to match the evaluation scale, which bottoms at 1, not 0. So, a very high evaluation average of 4.8 would result in an index score of 95 ($4.8 - 1/4 \times 100$), and a moderate to low average of 3.2 results in an index of 55 ($3.2 - 1/4 \times 100$). The other calculations were variations on this theme and always resulted in a number from 0-100.

There are weaknesses in amalgamating the data in this manner, notably the difference in the questions across schools, and also the psychological differences associated with the different numerical scales (i.e., students likely react differently to a form that places 1 as the highest score, rather than 5, and also probably treat a 7-point scale differently from a 4-point scale). Nevertheless, generating the index allowed for a mass correlation that covers the entire tenured and tenure-track faculty at 19 schools, a substantial advantage over 19 separate correlations. The correlation for each individual school, however, is included later for clarity's sake.

I also am aware that using teaching evaluations as a proxy for teaching effectiveness is somewhat controversial. There are studies, both within law schools and higher education in general, that show that teaching evaluations have biases, including biases based on race,¹⁷ gender,¹⁸ and even physical attractiveness.¹⁹ Other studies have shown that student teaching evaluations are positively correlated with other measures of teaching effectiveness, including peer reviews, suggesting at least that student measures track other alternative measures.²⁰

¹⁶ Also, based on an informal review of the data there was not generally a very big difference between student evaluations in lecture classes and seminars. The highest ranking professors had strong evaluations across the board, and lower-ranked professors were generally lower-ranked in all class sizes.

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 & L. 53, 165-200 (1999).

¹⁸ See, e.g., Christene H. Farley, *Confronting Expectations: Women in the Legal Academy*, 8 YALE J.L. & FEMINISM 333, 337-43 (1996).

¹⁹ See Mark T. O'Reilly, *Relationship of Physical Attractiveness to Students' Ratings of Teaching Effectiveness*, 51 J. DENTAL ED. 600, 600-02 (1987).

²⁰ See James A. Kultk, *Student Ratings: Validity, Utility, and Controversy*, and Michael Theall & Jennifer Franklin, *Looking for Bias in All the Wrong Places: A Search for Truth or a Witch Hunt in Student Ratings of Instruction?*, in THE STUDENT RATINGS DEBATE: ARE THEY VALID? HOW CAN WE BEST USE THEM? 9-26, 45-57 (Michael Theall, et al. eds., 2001); Herbert W. Marsh, *Students' Evaluations of University*

Many law faculty members nevertheless have argued to me that teaching evaluations are little more than a popularity contest. Some even have argued that teaching effectiveness is inversely correlated with teaching evaluations, since students tend to highly rank “easy professors” of little substance, while ranking those professors who challenge them comparatively lower.

For better or worse, teacher evaluations are the only viable way to measure teaching effectiveness for a study of this breadth. My other choices were exceedingly unpalatable: 1) attempt to gather peer evaluation data, which is rarely if ever expressed numerically, and would also almost certainly not be provided by the host institutions; or 2) use some type of personal subjective measure of teaching effectiveness, potentially requiring me to personally visit classes and make my own determination on teaching effectiveness.

Given that the teaching evaluation data are all that realistically could be gathered, the remaining question is whether the data are so unreliable that the data should be disregarded altogether. Since most institutions of higher learning bother to gather and compile the data, and since the data are regularly used in tenure, hiring, and promotion decisions, I decided the data had sufficient indicia of reliability for my purposes.²¹ Moreover, I note that almost all of the non-law studies of teaching effectiveness have used student teaching evaluations.²²

Lastly, on a personal note, the general tenor of the critique of law school teaching evaluations strikes me as extremely paternalistic and unfair. By the time law students fill out their student evaluation forms they have experienced at least 17 years of schooling, and frequently more. They are also poised to take on the responsibility of representing actual clients, appearing in court, and practicing law. Arguing that these same students are incapable of discerning good teaching from bad, or will evaluate professors in a popularity contest or based on the ease of the class, seems like a gross underestimation of their abilities and percipience.

C. What Was Studied – Scholarly Productivity

I also decided to measure the scholarly productivity of the same group of professors over the same four year period, 2000-03. After consultations with various law professors I settled on two different measures, one that emphasized purely “scholarly” activity (i.e., scholarly books and law review articles) and one that emphasized publications more directly focused on the practice of law (i.e., treatises, casebooks, and practitioner articles in bar journals).²³ I had originally planned on using only the “scholarly” measure and

Teaching: Research Findings, Methodological Issues, and Directions for Further Research, 11 INT’L J. ED. RES. 253, 263-336 (1987); DEREK BOK, *UNIVERSITIES IN THE MARKETPLACE* (2003)

²¹ Consider, for example, the sources cited *supra* note ___ or Herbert W. Marsh, *Students’ Evaluations of Teaching: Dimensionality, Reliability, Validity, Potential Biases and Utility*, 76 J. ED. PSYCH. 707 (1984).

²² See, e.g., Hattie & Marsh, *supra* note ___.

²³ Depending on your point of view either or both of these labels (“scholarly publishing” or “practice-oriented publishing”) may seem pejorative or insulting. No value judgment is intended; I use these labels only for ease of reference.

received strong objections from some faculty who felt that it discounted the importance of casebooks, treatises and the like. Some also commented that the study overly weighted scholarly publications and that those publications are the least likely to have a positive effect on teaching. In response to these criticisms I developed the second measure.

For both measures I gathered publication data for each individual professor. I followed the same procedure for each professor. I started with the law school’s website. Most of the 19 schools had at least some listing of each professor’s recent publications, and I used those listings as a baseline. Some schools even included a full CV for each faculty member, which was ideal. In order to make sure that each school and each professor was measured identically, I also checked each professor for publications on Westlaw and Amazon.com.

On Westlaw I checked the JLR database with the following search “au (first name /2 last name).” For example “au (benjamin /2 barton).” The “/2” was used to account for middle initials, and a /3 or /4 was used when middle names or initials made a larger search appropriate. If that search failed to produce any results I would check again for possible other names for publishing, like “au (ben! /2 barton)”. For Amazon.com I usually used as much of the professor’s name as possible to limit false positives.

If there were any sources listed on a professor’s resume or the law school’s website that were unavailable on Amazon or Westlaw I generally used a Google search to try to locate the publications. In short, I made every effort to find and verify each publication for all 623 professors.

From these three sources I generated a list of publications dated 2000-03 for each professor. I created a raw count of publications, with the only mediation being dividing any co-authored publication by the number of co-authors. This created a raw publication count for 2000-03.

I then separated the publications into different categories and assigned each category a point value (See Table A).

TABLE A – Productivity Points		
<u>Category</u>	<u>Scholarly Points</u>	<u>Practice-Oriented Points</u>
Scholarly Book	15	5
Chapter in a Scholarly Book	6	3
Top-20 or Peer-Reviewed Law Review Article (40+ pages) ²⁴	6	3
Law Review Article (40+ pages)	5	3

²⁴ As my measure of a top-20 law review I used the combined top-20 list in Colleen M. Cullen & Randall S. Kalberg, *Chicago-Kent Law Review Faculty Scholarship Survey*, 70 CHI.-KENT L. REV. 1445, 1454 (1995). I defined a publication as “peer-reviewed” if the editorial staff and the selection process were run by faculty members (whether faculty in law or another discipline).

Top-20 or Peer-Reviewed Law Review Essay (10-39 pages)	4	2
Law Review Essay (10-39 pages)	3	2
Treatise or Casebook	3	15
Practitioner Article or Chapter	1	5
Top-20 or Peer-Reviewed Publication under 10 pages	2	1
Law Review Publication under 10 Pages	1	1

I generated these categories and the points I assigned from an amalgam of sources and received wisdom. In creating the scholarly rankings I consulted a number of articles and essays offering scholarship advice to new law professors,²⁵ considered conversations with other professors, and used my understanding of the tenure process at Tennessee and other American law schools. I generated the practitioner-oriented rankings as a counter-point that emphasized publications aimed at practicing lawyers or law students. I created the point totals by essentially reversing the emphasis of the scholarly points.

There were, naturally, several additional wrinkles. If there were two or more authors for any publication, I divided the points by the number of authors. If a professor served as an editor for any of the above publications, she received half credit for that type of publication.

Another problem was how to count treatises or casebooks that were updated, but not originally published, during the period (2000-03). I decided to count any update for the full amount of points. I did this for two reasons. First, the sheer difficulty of following an area of law sufficiently to update a treatise or a casebook is hard to distinguish from the effort necessary to draft the first version.²⁶ I decided to err on the side of over-inclusion rather than under-inclusion. Second, the main argument of the treatise writers for separate productivity rankings was the potential positive teaching effect of being fully up-to-date in an area of law. Since an update within the period should reflect a professor who is up-to-date, I wanted to count updates fully.

It is worth noting the items I choose not to count. I did not count op-ed pieces or newspaper articles. I did not include briefs or other papers filed in a lawsuit. I chose not to count these items because they would be impossible to find independently, and I did not want to only include them depending on whether a faculty member or school decided to include them on its website or CV.

I also did not include faculty blogs. I chose not to count blogs for two reasons. First, when I asked Glenn Reynolds of Instapundit.com and Eugene Volokh of Volokh.com (two of the most influential law professor bloggers) they told me they would not count blogging in a ranking of scholarly productivity. Because they are leading lights in the area I took their advice seriously. Second, the difficulty of measuring the length and breadth of blogs for a scholarship ranking was too daunting to contemplate. Some law

²⁵ See, e.g., William R. Slomanson, *Legal Scholarship Blueprint*, 50 J. LEG. ED. 431 (2000).

²⁶ This is, of course, more true in some areas than others. For example, an update on a treatise on the inheritance tax likely had to be done almost from scratch during the period, while other areas may have changed little.

professors blog irregularly and some blog hourly. Some law professors only blog about narrow legal issues, and some blog about their cats. Lastly, blogs were much less common or accepted during the period studied. In sum, drawing distinctions among various law professor blogs seemed quite subjective and difficult, so I chose not to.

There are obvious problems with any ranking of scholarly productivity, and mine is no exception. On the one hand, it would be preferable to make as few subjective determinations as possible, and assigning any weight to faculty publications carries a number of value judgments. In this vein, some of the non-law school studies measure the raw number of publications or simply count the total number of published pages. I decided not to rely solely on this approach because there is a greater variation among law school publications than in many other disciplines, and I thought such a basic count on its own would not actually reflect professorial productivity. Nevertheless, the raw count is included as a bulwark against criticism of either weighted measure: if anyone objects to the points assigned she can always refer to the raw count.

In theory I or my research assistants also could have read all of the relevant publications and assigned points based on our own criteria/opinion.²⁷ Instead, I decided to divide the publications into categories and assign points as a compromise. This approach sufficiently captures the differences in the types of legal publications, but is not so subjective as to call the entire process into question.

D. What Was Studied – Scholarly Influence

Once I had completed my two categories of scholarly productivity a new layer of criticism arose. Most of the legal scholarship on the “value” of scholarship has been based in citation studies, not studies of raw productivity.²⁸ The proponents of citation studies argue that a correlation study that only measures productivity misses the entire point: the proper study is between scholarly influence and teaching, since scholarly influence is a much truer measure of publication quality.

Thus, I chose to measure scholarly influence and correlate those results against teaching effectiveness as well. I decided to do two measures of scholarly influence: one a lifetime total of “citations” in the Westlaw JLR database, and one per-year listing of citations, since a lifetime citation study may unfairly benefit longevity.

I gathered these two measures by more-or-less following Brian Leiter’s well-known citation study methodology.²⁹ First, I did a search in Westlaw’s JLR database for the

²⁷ When contemplating this option I am reminded of the scene in the movie *The Shining* where the Jack Nicholson character sits at a type-writer all day and types “All work and no play makes Jack a Dull Boy” continuously for days on end. See *THE SHINING* (Warner Brothers 1980).

²⁸ See, e.g., Theodore Eisenberg & Martin T. Wells, *Rankings and Explaining the Scholarly Impact of Law Schools*, 27 J. LEG. STUD. 369 (2000); James Leonard, *Seen’ the Cites: A Guided Tour of Citation Patterns in Recent American Law Review Articles*, 34 St. Louis U. Urban L. J. 181 (1990).

²⁹ See Brian Leiter, *Measuring the Academic Distinction of Law Faculties*, 29 J. LEG. STUD. 451 (2000); Brian Leiter, *Faculty Quality Based on Scholarly Impact*, at http://www.leiterrankings.com/faculty/2005faculty_inpact_cites.shtml (last visited September 11, 2006).

professor's name. I searched for "first name /2 last name". For example, my search was "Benjamin /2 Barton". I again expanded the search for multiple middle initials or names, and considered alternate searches if a first search resulted in few or no cites. So, if a Benjamin /2 Barton search came up with no citations, I would try "Ben! /2 Barton" to catch nicknames. Because this information is time sensitive I did the entire study of scholarly influence in a short time frame (March 23-27, 2006). If any additional cites were added during these few days it should not have been enough to make a substantial difference.

I then counted the raw number of references returned. If the professor had a common name (like mine) I would scan the first 20 references for false positives. If there was a false positive I would extend the search to 40 total cites. In order to remove the false positives I would then divide the number of false positives by 40, and multiply that percentage by the entire number of citations. If the false positive count overwhelmed the true references, I would try another search, rather than risk an inaccurate count. For example, a search for "John /2 White" brings back so many different John Whites that the entire count could be inaccurate, regardless of the false or true references in the first 40. In that case I would try "John A. White", adding the middle initial to narrow the search down.

Please note that this Westlaw search counts each professor's scholarly influence through March, 2006. I decided to stretch the time frame for scholarly influence beyond 2003 because measuring only citations in 2000-03 would actually fail to capture the influence of the publications being written during the timeframe of the study because of the lag between writing, publishing, and eventual citation.

I also counted the number of years each faculty member had been in full-time tenure-track law teaching. I then divided the total number of references by the number of years to get a measure of references-per-year. I created the references-per-year number so that longevity in teaching was not ignored in the citation study and so that there was a measure that captured "influence-per-year," rather than just a career's worth of citations.

There are a few notable problems with this measure. First, as Brian Leiter has also recognized, it actually captures more than just citations. It captures "star cite" references, and self-citations, both of which might be disallowed from a true "citation-only" study. Because I am measuring "influence," and my rough eye-balling of the references show that few law professors have an unusual ratio of citations to star cite references, I chose not to weed through the raw numbers to eliminate those references.³⁰ Further, insofar as I am attempting to measure "influence," a star cite reference hopefully shows that the named professor had at least some influence on the publication at issue.

³⁰ This also would have added an entire layer of potential errors and bias in deciding what was a "real" citation and what was not. It would have also been impossible to accomplish over a short period of time (since some professors had more than 1000 references in Westlaw), and that would have introduced the need to adjust for time in the study, adding another possible source of error.

Second, I did not do a corresponding study of judicial citations to professors. I did not do so because court citations are much harder to find on Westlaw and much rarer, and I worried about my ability to accurately count those citations.³¹ The practice-oriented side of this measure is thus missing, and I recognize the weakness of this asymmetry.

Nevertheless, I thought adding two influence measures to the productivity measures was a major addition to the study, since it covered almost any conceivable measure of publication value, and it eliminated some of the weaknesses of the productivity measures. For example, the productivity measure only covers 2000-03, and some very productive scholars across their careers may have slowed over those four years for whatever reason. If scholarly productivity actually affects teaching, it is worthwhile to try to measure a lifetime effect as well as a set time period.

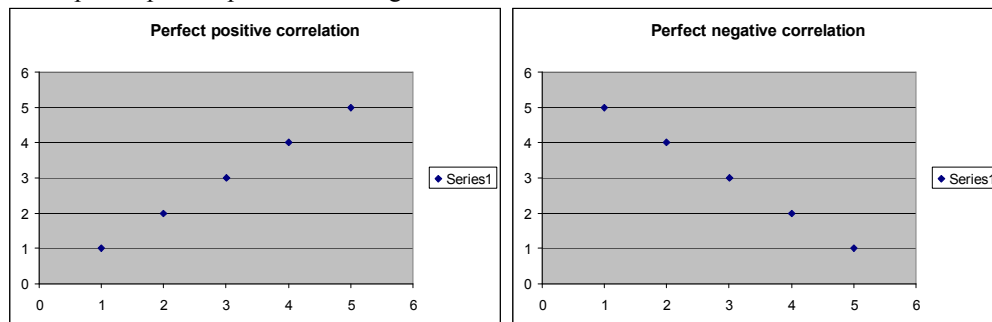
III. RESULTS

The short version: none of the five different measures of scholarly productivity or scholarly influence correlate with teaching effectiveness. To reach this conclusion, I gathered all the numbers on a single spread sheet and used Stata 9.0 to calculate a Pearson's coefficient, as well as a p-value to measure statistical significance.

The Pearson's coefficient calculates how strongly two variables correlate. A perfect positive correlation coefficient would be +1. A perfect negative correlation is a -1.³² A negative correlation coefficient means that as one number grows larger, the other grows smaller; the two variables move in opposite directions. A positive coefficient means that as one grows larger the other does too; the two variables move in the same direction. As a general rule of thumb, there are five categories of correlation coefficients: strong positive coefficients (+.7 to 1), strong negative coefficients (-.7 to 1), weak positive coefficients (+.4 to .7), weak negative coefficients (-.4 to .7), and no correlation, or very weak correlations (-.3 to .3).³³ A coefficient of 0 is a perfect finding of no correlation.

³¹ Many law professors have relatively common names that occur frequently as names of plaintiffs or defendants, creating a massive false positive problem. Further, the courts are not as regular as law reviews in their citation style, and many times do not provide the professor's first name, making it difficult to gather an accurate citation count.

³² Graphs of perfect positive and negative correlation look like this:



³³ See DAVID COPE, *FUNDAMENTALS OF STATISTICAL ANALYSIS* 63 (2005).

The p-value measures the likelihood that the null hypothesis is correct. In this case the null hypothesis is that there is no correlation between the two variables studied. When the p-value rises above either .05 or .10 the coefficient results are generally not considered statistically significant, because the null hypothesis of no correlation is too likely to draw any firm conclusions from any finding of correlation. In the results below the correlation coefficients are quite small (basically indistinguishable from zero) and the p-values are quite high, indicating that the null hypothesis of no correlation is very likely correct.

Here are the findings for a correlation between the teaching evaluation index and each of the five scholarship measures:

	<u>Correlation Coefficient</u>	<u>P-Value</u>
Scholarly Productivity:	.0029	.9417
Practice-Oriented Productivity:	.0366	.3614
Raw Publications Count:	.0229	.5677
Total Citations:	-.0139	.7298
Citations per Year:	.0224	.5770

N= 623

This is an extremely strong finding of no correlation, since each coefficient is so close to zero. Further, even the slight positive and negative correlations that were found are statistically insignificant by a great margin. In sum, this study of 623 professors shows that regardless of how you measure scholarly productivity or scholarly influence there is no correlation with teaching effectiveness.

A review of the individual correlation coefficients and p-values for each participating school further underscores this finding (see Table B):

TABLE B	Scholarly	P Value	Practice	P Value	Total Pubs	P Value	Cites Total	P Value	Cites Per Year	P-Value
Colorado	0.0594	0.7551	0.1288	0.4976	0.0657	0.7303	0.1065	0.5753	0.1375	0.4687
Connecticut	0.078	0.6714	-0.0214	0.9076	0.01	0.9566	0.1538	0.4006	0.1094	0.5511
Cumberland	0.2507	0.2267	0.1315	0.531	0.1693	0.4186	0.16	0.445	0.1199	0.5682
Florida	0.059	0.721	-0.1197	0.4679	0.0351	0.8318	-0.087	0.5982	0.0806	0.6258
Iowa	-0.1392	0.4046	0.042	0.8022	-0.0574	0.732	0.1136	0.4971	0.1289	0.4405
Lewis & Clark	-0.1311	0.498	0.0709	0.7149	0.0531	0.7843	0.1247	0.5192	0.0868	0.6543
Michigan	-0.0258	0.883	0.0538	0.7587	0.0118	0.9463	0.1525	0.3819	0.1867	0.2827
UND	-0.1184	0.7446	0.1268	0.7271	0.0056	0.9878	0.3319	0.3487	0.1834	0.6121
Northwestern	0.003	0.9851	-0.0023	0.9885	-0.005	0.9752	-0.0843	0.6004	-0.1391	0.3859
Ohio State	0.3232	0.0582+	0.3494	0.0396*	0.3087	0.0712+	0.082	0.6394	0.1339	0.4433
Penn State	0.2914	0.1671	0.2928	0.165	0.2213	0.2988	0.1025	0.6336	0.3435	0.1003+
Southwestern	-0.1409	0.3987	-0.1282	0.4429	-0.1987	0.2317	-0.1735	0.2975	-0.0482	0.7738
St. John's	0.2993	0.072+	0.251	0.1339	0.2605	0.1194	0.0851	0.6165	0.2716	0.1039
Tennessee	-0.1515	0.3848	0.219	0.2063	-0.0079	0.9639	0.0103	0.953	0.0656	0.7082
Texas Tech	-0.0371	0.8378	0.0144	0.9364	-0.0427	0.8137	0.0754	0.6767	-0.1535	0.3937

Toledo	0.103	0.632	-0.238	0.2628	0.0492	0.8195	0.2086	0.328	0.1719	0.4218
UCLA	0.0877	0.5127	0.2583	0.0503*	0.1291	0.3342	-0.0772	0.5646	-0.0167	0.9009
Villanova	0.1432	0.4191	-0.086	0.6285	0.1115	0.5302	0.1264	0.4761	0.3912	0.0222*
Wayne State	-0.0993	0.6293	-0.3523	0.0775+	-0.171	0.4035	-0.1436	0.4841	-0.1819	0.3737

+ = significant at .10

* = significant at .05

A. What the Results Mean

The overall correlation study reaches an extremely strong finding of no correlation, with coefficients as close to zero as any such study can reach, and very high p-values. Further, the separate coefficients for each school also strongly support a finding of no correlation. Out of 95 separate correlations, there are only 3 correlations that are significant at .05, and only 8 that are significant at .10. The strength of this conclusion is buttressed by the different measures of research productivity and influence used. Based on the five different measures it is difficult to argue that this study does not capture research productivity/influence in some form or another.

It is interesting to note that Ohio State has three of the eight statistically significant correlations, suggesting that the most prolific members of their faculty for the period 2000-03 were also the highest rated teachers. It is also interesting to note that the “total citations” measure has no statistically significant correlations, suggesting that it is the least likely determinant of high teaching evaluations.

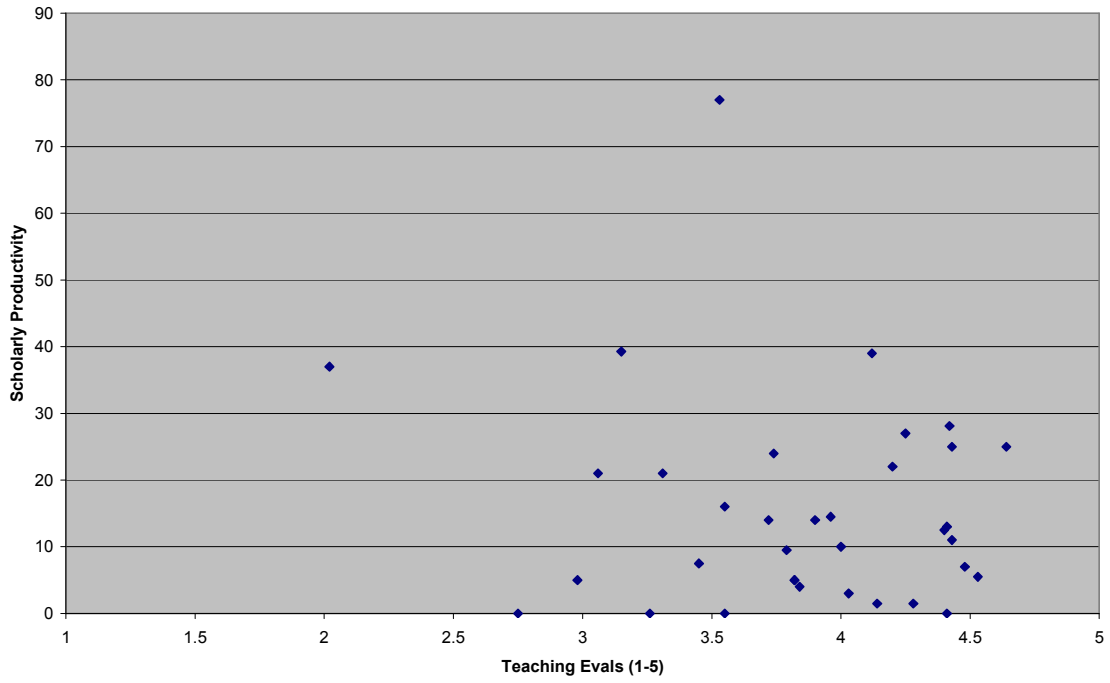
The results are thus quite clear: regardless of whether the professors are amalgamated onto a single correlation or the schools are considered separately, and regardless of what measure of research productivity is used, there is no correlation between teaching effectiveness and scholarly productivity. A comparison of the many different types of law schools on Table B and the general uniformity of results is also striking. The correlation coefficient is similar (and very weak) regardless of whether the schools are public or private, larger or smaller, stronger or weaker in academic reputation, stronger or weaker in overall scholarly productivity, and regardless of whether the school is in the east, west, south or north of the U.S.

A finding of no correlation means that an American law professor is equally likely to fall into one of the following four categories:

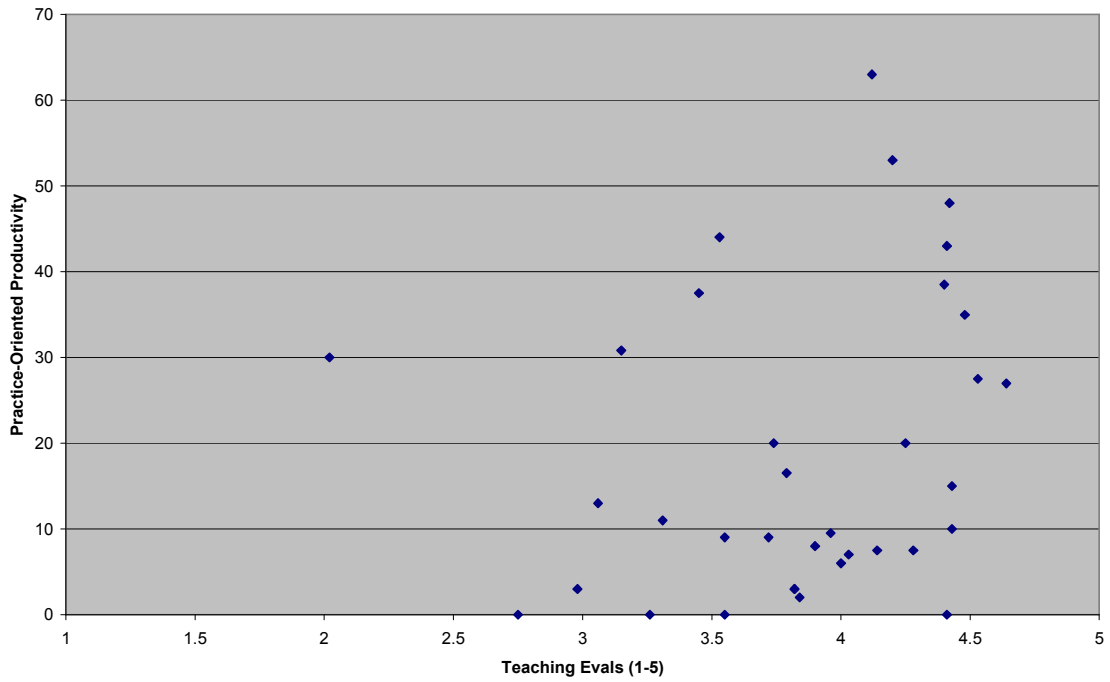
- 1) strong teaching evaluations and strong scholarly productivity;
- 2) weak teaching evaluations and weak scholarly productivity;
- 3) strong teaching evaluations and weak scholarly productivity; or
- 4) weak teaching evaluations and strong scholarly productivity.

As a visual aid to how unstructured the data is, consider the following five charts (Figure C) from the University of Tennessee data:

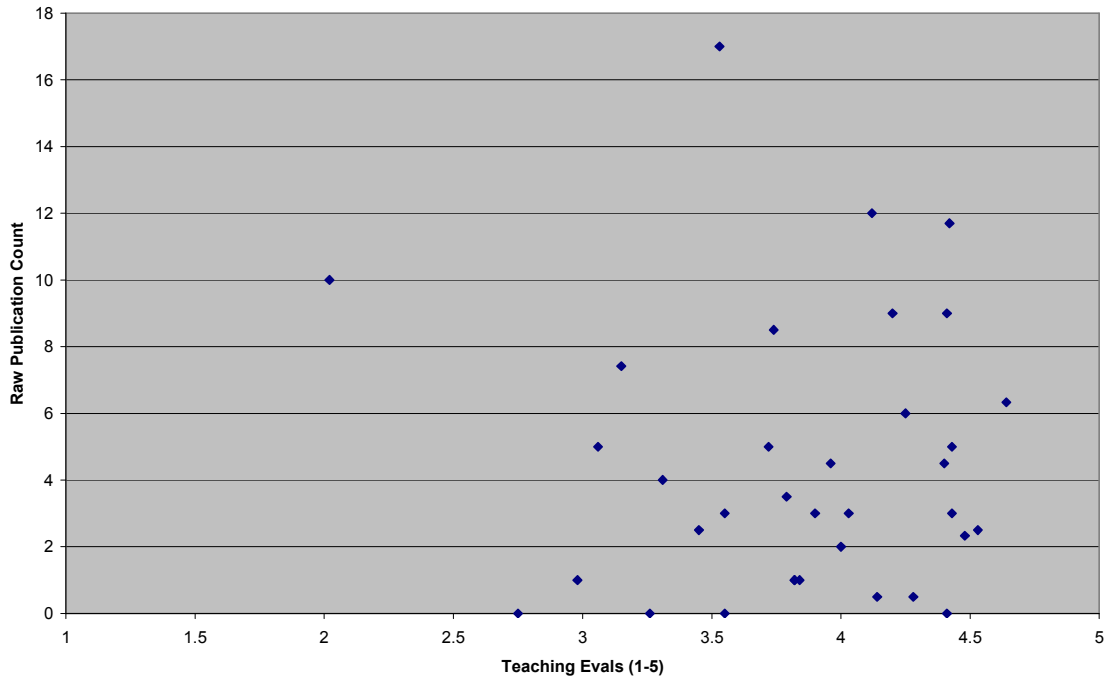
Tennessee Scholarly Productivity



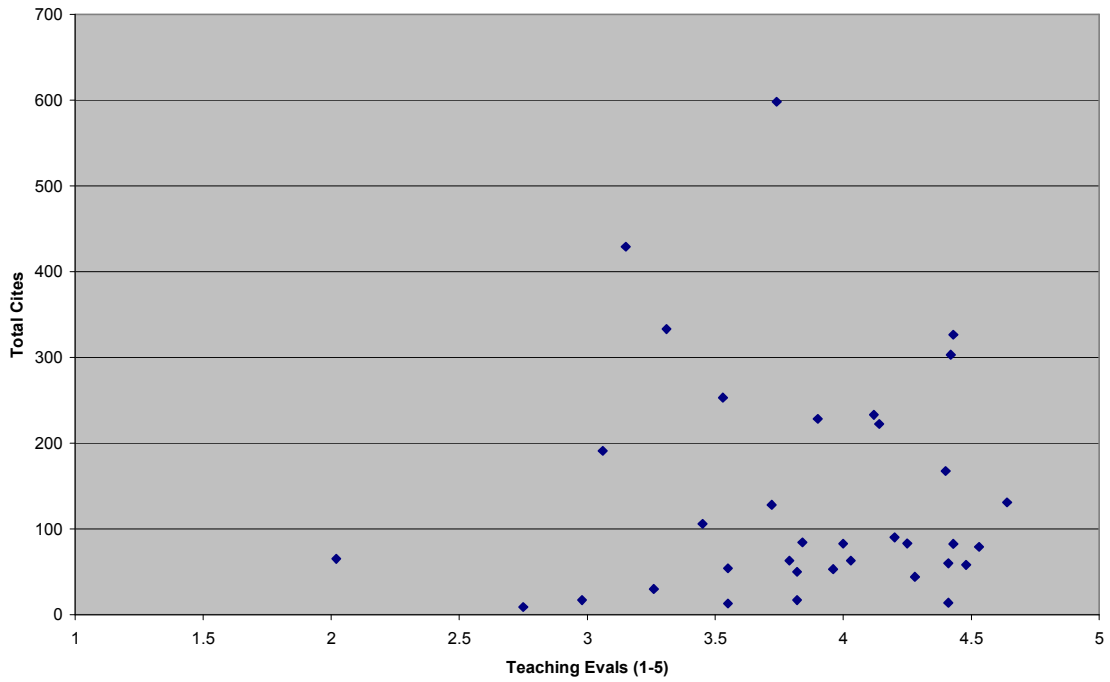
Tennessee Practice-Oriented Productivity



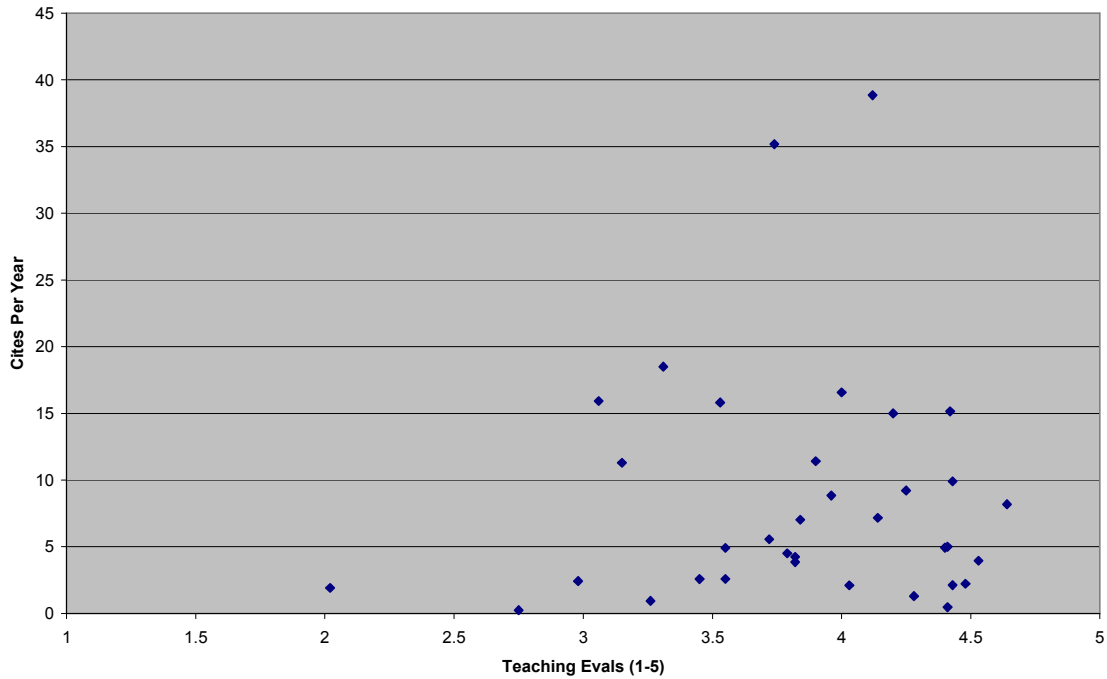
Tennessee Raw Publications



Tennessee Total Cites



Tennessee Cites Per Year



A review of these four charts shows how spread out the data is, and how impossible it would be to map the results in any kind of straight line. In sum, none of these five measures of research productivity correlate with teaching evaluations.

B. The Counter-Intuitive Nature of this Finding

In describing this study to others, I have found three general predictions as to my results. A slight majority have predicted that scholarly productivity and teaching effectiveness will be positively correlated. They reason that the process of researching and writing scholarship should have multiple benefits for classroom teaching: the professor is more likely to be up-to-date in her knowledge of a particular subject area, and her thinking on the topic should be polished and honed in the process of drafting scholarship. The proponents of practice-oriented scholarship were particularly vociferous on this point, and it intuitively makes sense that the professors who focus on writing casebooks, treatises, and practitioner articles would have many advantages in teaching those same materials. In short, many thought that excellence in teaching and scholarship would be positively correlated.

There was a second, smaller group, which postulated a negative correlation. Since faculty members have limited time for teaching and scholarship, some guessed that the extra time spent on producing scholarship would lessen teaching effectiveness. In this zero-sum projection, the professors who spent the bulk of their energy on publishing would naturally suffer as teachers. Last, some had no idea how the numbers would correlate.

Unless a person had read the literature outside of law schools that shows a general finding of no correlation in other disciplines, no one guessed that there would simply be no correlation between the numbers. Statistically speaking it is unwise to conflate correlation and causation, and sometimes even correlation numbers can be misleading, if there is a factor outside of those studied that is driving the numbers. Nevertheless, a finding of no correlation is generally considered strong evidence that two factors are not significantly affecting each other one way or another.

IV. CONCLUSION

As a matter of law school policy this finding is both helpful and puzzling. In fact, throughout the process of gathering and analyzing the data I have found it to be something of a Rorschach test -- people tend to read in their own preferences. For example, some faculty members that think that there is too much emphasis on publishing in legal academia have responded: "Good. At last we'll learn that hiring and promoting solely on the basis of scholarship does nothing to help/teach our students." Some faculty members that are more pro-scholarship have said: "Excellent. At last the myth that scholarship and teaching are at odds has been debunked."

Without joining either side of this debate I will add the following to those reactions. First, if it is true that with the advent of rankings and increased faculty competition law schools have begun to focus more on scholarship in hiring and promoting, that practice may not do much to further a school's teaching mission.

Second, while hiring and promoting on the basis of scholarship may not do much to further the schools teaching mission, the study also suggests that scholarly productivity is not at odds with teaching. To the contrary, it seems to have little effect. So, insofar as law schools are pushing faculty to increase scholarly productivity it does not appear to be having a deleterious effect on teaching.

Lastly, it is worth thinking about why the AALS and ABA accreditation standards basically require every law school in America to have a sizable teaching faculty that also engages in scholarship,³⁴ if scholarship and teaching effectiveness are unrelated. There are other reasons why the AALS and ABA might require scholarly productivity. It might be considered a service to the bench, bar, or society at large, or it might help the students or the school as a whole in a way uncaptured by teaching evaluations. Nevertheless, if these other justifications underlie those rules, the ABA and AALS should say so.³⁵

³⁴ See ABA, ABA Standards for Approval of Law Schools, Standard 404; AALS, AALS By-Law Section 6-4(a).

³⁵ Note that I am not suggesting that scholarship is not separately valuable, or that law schools should not have faculties full of productive scholars. I am only saying that if that is a (or the) primary goal of American law schools it should be defended aside from the effect on the teaching of students, since this study shows that there is no demonstrable effect.

As a sidenote, consider the findings of Marsh and Hattie that extra time spent on research activities results in increased research productivity, but that extra time spent on teaching does not improve teaching evaluations.³⁶ This may suggest that teaching (at least as measured by teaching evaluations) is more of a matter of innate ability or personality style than research productivity. When combined with this study's strong finding of no correlation, it is worth considering whether the skills and attributes that underlie research productivity and teaching evaluations are similar or even connected. If these skills are dissimilar, what (if anything) can American law schools do to try to hire and promote professors that are good researchers and teachers? Is there anything that can be done structurally to encourage excellence and/or progress in both measures?

In sum, this study answers a very basic question about the relationship between scholarship and teaching at American law schools. However, many fundamental questions about the nature and mission of law schools and the legal professoriate remain unanswered.

³⁶ See John A.C. Hattie & Herbert W. Marsh, *The Relation between Research Productivity and Teaching Effectiveness – Complementary, Antagonistic, or Independent Constructs?*, 73 J. HIGHER ED. 603, 611-13 (2002).