

University of Tennessee College of Law

Legal Scholarship Repository: A Service of the Joel A. Katz Law Library

UTK Law Faculty Publications

9-2008

Is There a Correlation between Law Professor Publication Counts, Law Review Citation Counts, and Teaching Evaluations? An Empirical Study

Benjamin H. Barton

Follow this and additional works at: https://ir.law.utk.edu/utklaw_facpubs



Part of the Law Commons



Legal Studies Research Paper Series

**Research Paper #81
November 2009**

**Is There a Correlation between Law Professor
Publication Counts, Law Review Citation Counts, and
Teaching Evaluations?
An Empirical Study**

Benjamin Barton

This paper can be downloaded without charge
from the Social Science Research Network Electronic library at:
<http://ssrn.com/abstract=913421>

Is There a Correlation Between Law Professor Publication Counts, Law Review Citation Counts, and Teaching Evaluations? An Empirical Study

Benjamin Barton¹

ABSTRACT

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 correlates these data against five different measures of research productivity/scholarly influence.

The results are counter-intuitive: there is either no correlation or a slight positive correlation between teaching effectiveness and any of the five measures of research productivity. Given the breadth of the study, this finding is quite robust. These findings are sure to spark heated debates among law faculties and likely require some soul-searching about the interaction between the two most important functions of American law schools.

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 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.

Is There a Correlation Between Law Professor Publication Counts, Law Review Citation Counts, and Teaching Evaluations? An Empirical Study

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 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 faculty that neglects scholarship that is actually harming students. This debate has also been echoed in various law review articles (Korobkin, 1998 (pro-scholarship) and Scordato, 1990 (pro-teaching)).

On the one hand, it seems likely that working hard on scholarship should have a 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.

The question itself is likely impossible to answer definitively, especially since there is little agreement about how to measure either the quality of teaching or the quality/impact of legal scholarship. This study does not try to answer these questions once and for all, but does the best it can with the available data. I gathered three different types of data from the tenured or tenure-track faculties at nineteen American law schools. I gathered four years of teaching evaluations. I also did a publication count across those same four years. Lastly, I gathered law review citation data. I used the publication count and citation data to generate a total of five different measures of scholarly output, and correlated each measure against teaching evaluations. The study found that there is either no correlation or a slight positive correlation between teaching evaluations and publication count or citation counts.

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 (Hattie and Marsh 1996), John Braxton’s synthesis (Braxton 1996), and Kenneth Feldman’s earlier collection of studies (Feldman 1987). 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. Because 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.

Hattie and Marsh sought to avoid many of these problems in their more recent study of 182 Australian university professors (2002). Their thorough study used multiple measures of research productivity, and also made a strong finding of no correlation. I followed many of the methodological suggestions of Hattie and Marsh in designing 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 (Merritt 1998). 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. Furthermore, using teaching awards as a proxy for teaching effectiveness is somewhat problematic, because these awards are handled differently at every institution.

James Lindgren and Allison Nagelberg conducted a second study (Lindgren and Nagelberg 1998). 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 (Eisenberg and Wells 1998). 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.

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 (Lindgren and Nagelberg 1998). 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 evaluations and scholarly productivity in American law schools. I planned to gather teaching evaluation

data for a four year period (2000-2003) from twenty American law schools, and 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.²

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 a good 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 schools (ABA-Approved Law Schools 2006) 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.

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

² A copy of the original prospectus can be found at <http://www.law.utk.edu/FACULTY/BartonProp.PDF>.

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 therefore could have skewed the analysis.

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 Evaluation Data

For each school I chose the question on the teaching evaluation sheet that most closely measured teaching effectiveness. For example, The University of Tennessee College of Law's form 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.³

In making the individual, school-level correlations I used the teaching evaluation data as-is and then correlated it against the various measures of productivity and influence. In order to make a mass correlation amongst all the schools I made two transformations.

The first was a linear transformation of each school's teaching evaluations. I converted each professor's average teaching evaluation score into a teaching evaluation index that ranged from 0-100. 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 0-100.⁴

³ Wherever possible I used a per student response average. Four of the schools supplied class level averages only, and one provided only averages by professor. I recognize that averaging the responses discards the richness of the student level data I have for some of the other schools. This data will be utilized in the later multivariate analysis.

⁴ 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.

Second, I generated a z-score for each professor within their school's teaching evaluations. The z-score calculated the distance from each school's mean teaching evaluation scores in standard deviation units. These z-scores were then amalgamated and used for various mass correlations.⁵

There are weaknesses in amalgamating the data in either of these ways, and I recognize that some readers may discount the mass correlations because of those weaknesses. In particular, there are the different questions considered across the schools, and 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). Moreover, most of the students at any particular school will have had no exposure to the teaching of professors at any other school considered, so the rankings are likely to be relative within each law school.

Nevertheless, generating the two indexes allowed for mass correlations that cover the entire tenured and tenure-track faculty at 19 schools, a substantial advantage over the 19 separate correlations. Further, some of the concerns about these transformations are blunted by the uniformity of the eventual results: none of the individual schools showed a strong correlation between teaching evaluations and any of the measures of scholarship, and the mass correlations reached a similar result.

I also am aware that using teaching evaluations at all is somewhat controversial. There are studies, both within law schools and higher education in general, that claim to establish that teaching evaluations have biases, including biases based on race (Smith 1999), gender (Farley 1996), and even physical attractiveness (O'Reilly 1987). The methodologies of these, and other studies finding bias, are open to criticism, however. (Marsh 2007).

Many law faculty members 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.

I have three responses to these objections that I offer in ascending order of stridence. First, teacher evaluations are the only viable way to even attempt to measure teaching effectiveness for this type of study.⁶

⁵ For a description of what z-scores are and how they are calculated, please see Newton & Rudenstam (1999). I used a program called "zscore" in Stata 9.1 to calculate the zscores.

⁶ 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. Moreover, I note that almost all of the non-law studies of teaching effectiveness have used student teaching evaluations (see, for example, Hattie & Marsh 1996).

Second, even the most vociferous critic of teaching evaluations has to admit they measure *something*. Notably, they measure what the students who responded on a particular day thought of the professor's performance in that class over the course of the semester. Even if critics contend that teaching evaluations utterly fail to capture teaching effectiveness, they can limit themselves to the question of whether teaching evaluations (whatever they measure) correlate with scholarly productivity.

Third, studies have shown that student teaching evaluations are valid and positively correlated with other measures of teaching effectiveness, including peer reviews. These studies suggest that, at a minimum, teaching evaluations track other evaluations of teacher effectiveness. (Bok 2003; Marsh and Roche 2000; Marsh 1987 and 1984). Marsh's recent overview of the validity and multidimensionality of student evaluations offers a particularly persuasive and robust defense. (Marsh 2007).

Lastly, the critics of teaching evaluations seem to assume that law students are incapable of accurately measuring the teaching they receive in law school. As a relatively recent graduate of law school I find that argument patronizing at best and insulting at worst. Law students have sat through a minimum of sixteen years of organized instruction before they rate their first law school class, and it defies common sense to say that they have learned so little about discerning good teaching from bad that they cannot accurately rank a professor's teaching effectiveness on a five point scale.

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 three similar measures, one that simply counted each professor's publications during 2000-03, one that emphasized purely "scholarly" activity (i.e., scholarly books and articles) and one that emphasized publications more directly focused on the practice of law (i.e., treatises, casebooks, and practitioner articles in bar journals).⁷

For all the measures I gathered publication data for each individual professor.⁸ From these three sources I generated a list of publications dated 2000-03 for each professor. I

⁷ 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.

⁸ 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.

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 two different point values: the relative value for scholarly purposes and the relative value for practical lawyering purposes (See Table 1).

TABLE 1 – 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) ⁹	6	3
Law Review Article (40+ pages)	5	3
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 (see, for example, Slomanson 2000), 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.¹⁰

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.

⁹ As my measure of a top-20 law review I used the combined top-20 list in Cullen and Kahlberg (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). Some commenters have worried that this study focused too narrowly on law review articles. To the contrary, scholarly publications in non-law academic journals were counted, and were almost invariably granted the bonus points for peer-review, making non-law publications at least as valuable, and often slightly more valuable, than law review publications.

¹⁰ 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. Second, the main argument of the treatise writers for separate, practice-oriented productivity rankings was the potential positive teaching

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 accurately 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. 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

I also decided to do a citation study of the same 623 professors. I did this for several reasons. First, I wanted to match the legal scholarship that has been based in citation studies, not studies of raw productivity (Eisenberg & Wells 2000; Leonard 1990). This addition also answers the criticism that a correlation study that measured only 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. Second, a citation measure covers a lifetime of work, so it should catch any scholars who were less productive during 2000-03 for whatever reason.

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 following Brian Leiter’s well-known citation study methodology (Leiter 2000). I describe the methodology in more detail in the footnotes, but I basically used Westlaw’s JLR database

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 because of the impossibility of measuring blog activity in 2000-03 and the great imbalance among different blog types and activity levels.

to generate a raw list of citations for each professor and then I created a “cites-per-year” measure by dividing the total number of cites by the years in full-time teaching.¹¹

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.¹² 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.

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.¹³ The practice-oriented side of this measure is thus missing, and I recognize the weakness of this asymmetry.

¹¹ The full methodology is as follows: First, I did a search in Westlaw’s JLR database for the 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.

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.

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 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.

¹² 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.

¹³ 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.

Lastly, Westlaw does not include many non-law publications, so professors who publish primarily in the journals of other disciplines or are cited primarily in those journals will be undercounted by this citation measure. This weakness is blunted by the earlier productivity measure, which explicitly includes non-law scholarly publications found via the home school web site, Amazon and Google. Moreover, to my knowledge there is not a non-law equivalent to Westlaw, so a broader citation count would be impossible. Lastly, whatever argument there is for a connection between scholarly productivity and teaching, it is hard to see how citation and influence in non-law journals would have much effect on law school teaching.

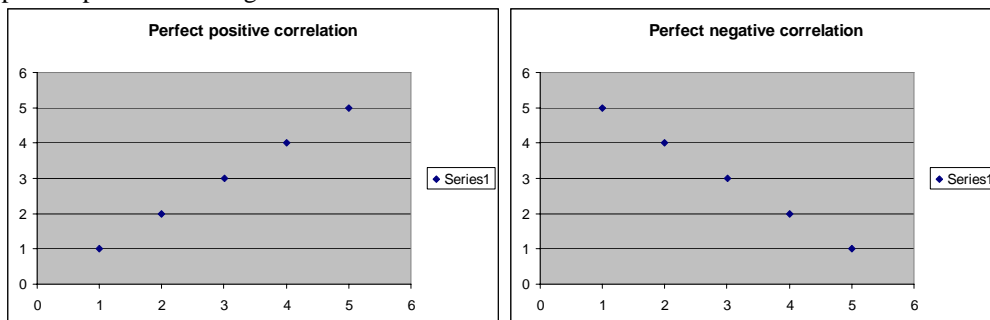
III. RESULTS

I gathered all the numbers on a single spread sheet and used Stata 9.1 to calculate three different sets of correlation coefficients. I used the raw numbers to create a Spearman rank order coefficient, as well as p-values to measure statistical significance. I used a Spearman correlation first because much of the data were not normally distributed, so a non-parametric correlation measure like the Spearman was more appropriate.

Next I adjusted as much of the data as I could to a normal distribution and then ran a series of Pearson product moment correlations on the adjusted numbers. The numbers were adjusted because the Pearson correlation is parametric and requires a normal distribution. Lastly, I ran a Pearson correlation on the non-normal, raw data as a companion to, and check on, the other two sets of correlations. All three of the correlation sets were largely consistent, which supports the overall findings and conclusions.

Both the Spearman and Pearson correlations generate a correlation coefficient that measures how strongly two variables correlate.¹⁴ The p-value measures the likelihood

¹⁴ A perfect positive correlation coefficient would be +1. A perfect negative correlation is a -1. Graphs of perfect positive and negative correlation look like this:



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

that the null hypothesis is correct, i.e. the p-value helps avoid type I errors.¹⁵ 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.

A second concern is type II errors.¹⁶ In order to avoid type II errors many researchers look to the power of their correlation. If the power is below .8, then the risk of a type II error is too high for the null hypothesis to be definitively accepted. The power of a correlation is calculated from the sample size and the relative value of the correlation coefficient sought. As correlations become stronger it is possible to detect them in smaller samples. As the correlations sought become weaker, the sample size must be bigger to reliably detect a correlation. Similarly, if you require a p-value of .05 or lower to find a correlation, the sample size required is larger than if a p-value of .1 (or higher) is required. In short, the sample size required depends on the size of the correlation the researcher is seeking, as well as the stringency of the p-values required.

The following chart shows the sample size necessary in order to surpass a power of .8, broken down by the size of the correlation coefficient sought and the p-value required:¹⁷

<u>Correlation Coefficient</u>	<u>P-Value</u>	<u>Sample Size Required</u>
.5	.1	18
.5	.05	23
.3	.1	50
.3	.05	67
.1	.1	453
.1	.05	619

As you will see below, the entire sample of professors (n=623) is large enough to meet the .8 power requirement, even when searching for correlations as weak as .1. Nevertheless, the sample sizes for the individual schools (which range from 10 to 58) are often too small to detect low levels of correlation. I break out the scores by school regardless, because it helps explicate and support the overall numbers. Readers should note that the number of professors at any individual schools (which are fixed and cannot be enlarged) may often be too small to fit a .8 power restriction, and read the results accordingly.

The results will be reported in the following order. First I will report the mass Spearman rank order correlations for all 623 law professors. I will then show the Spearman

correlations (-.3 to .3) (Cope 2005). Some social scientists consider any finding of correlation above .1 as a positive correlation. A coefficient of 0 is a perfect finding of no correlation.

¹⁵ A type I error is when a researcher finds a correlation when in fact the null hypothesis of no correlation is in fact true.

¹⁶ Type II errors occur when a researcher discards a finding of correlation in favor of the null hypothesis of no correlation in error.

¹⁷ These numbers were calculated using the SISA sample size calculation (SISA).

correlation coefficients broken out by school. Next I will report the mass and school-by-school Pearson correlations with all non-normal data transformed to normality wherever possible. Lastly, I will produce the unadjusted Pearson correlations as a comparison point to the earlier Spearman and transformed Pearson correlations.

A. Spearman Correlations

Here are the Spearman rank order correlation coefficients between the teaching evaluation index and the z-scores and each of the five scholarship measures:

	<u>Index Spearman</u>	<u>P-Value</u>	<u>Z score Spearman</u>	<u>P-Value</u>
Scholarly Productivity:	0.0627	0.1180	0.0866	0.0306
Practice-Oriented:	0.0730	0.0687	0.1301	0.0011
Raw Publications Count:	0.0607	0.1303	0.0903	0.0242
Total Citations:	0.0114	0.7761	0.0343	0.3932
Citations per Year:	0.0468	0.2433	0.0588	0.1429

N= 623

All of the coefficients but one are under .1, and many of the small correlations that are found are not statistically significant. The most that could be said of these coefficients is that the z-scores had a slight positive relation to the raw publication, scholarly productivity, and practice-oriented publication counts. The correlation coefficients of the index version of the teaching evaluations are basically indistinguishable from zero, and are not statistically significant at .05, although the small positive correlation found for practice-oriented scholarship is statistically significant at .10. Overall, these findings support the conclusion of no correlation, or at most a very small correlation with scholarship produced.

A review of the individual Spearman rank order correlation coefficients and p-values for each participating school further underscores this finding (see Table 2):

	Scholarly	P-Value	Practice	P-Value	Total Pubs	P-Value	Cites Total	P-Value	Cites Per Year	P-Value
Colorado (N=30)	0.1828	0.3336	0.2401	0.2012	0.2194	0.2440	0.2005	0.2880	0.2652	0.1567
Connecticut (N=32)	-0.0310	0.8661	0.0565	0.7586	0.0215	0.9068	0.1568	0.3914	0.1419	0.4384
Cumberland (N=25)	0.2850	0.1673	0.1807	0.3874	0.2438	0.2403	0.0943	0.6538	-0.0092	0.9651
Florida (N=39)	-0.1061	0.5205	-0.0931	0.5731	-0.1173	0.4769	0.0539	0.7444	0.0878	0.5952
Iowa (N=38)	-0.1074	0.5211	0.0732	0.6622	-0.0553	0.7418	-0.0959	0.5669	-0.0279	0.8679
Lewis & Clark (N=29)	0.0679	0.7264	0.1536	0.4262	0.1703	0.3772	0.2002	0.2978	0.1859	0.3344
Michigan (N=35)	0.2681	0.1194	0.1943	0.2635	0.2295	0.1848	0.1522	0.3826	0.1758	0.3123
UND (N=10)	-0.1506	0.6778	-0.0490	0.8932	-0.1077	0.7671	0.3211	0.3656	0.3761	0.2840
Northwestern (N=41)	-0.0069	0.9659	-0.0889	0.5804	-0.0232	0.8855	-0.1708	0.2856	-0.1100	0.4937
Ohio State (N=35)	0.3166	0.0639+	0.3401	0.0456*	0.3053	0.0745+	0.0780	0.6559	0.1143	0.5132
Penn State (N=24)	0.3856	0.0627+	0.4018	0.0517*	0.2339	0.2713	0.0661	0.7589	0.2022	0.3433
Southwestern (N=38)	-0.2173	0.1901	-0.2320	0.1610	-0.2635	0.1100	-0.2610	0.1135	-0.0830	0.6205
St. John's (N=37)	0.3618	0.0278*	0.2751	0.0994+	0.3644	0.0266*	0.1332	0.4319	0.2958	0.0754+

Tennessee (N=35)	0.0133	0.9394	0.2803	0.1029	0.0936	0.5927	0.0912	0.6025	0.0433	0.8050
Texas Tech (N=33)	-0.1047	0.5619	-0.0568	0.7534	-0.0780	0.6662	-0.0425	0.8145	-0.1915	0.2856
Toledo (N=24)	-0.0538	0.8030	-0.1885	0.3777	-0.0865	0.6877	0.1588	0.4586	0.1388	0.5178
UCLA (N=58)	0.1779	0.1815	0.3607	0.0054*	0.1791	0.1786	-0.0201	0.8812	-0.0515	0.7009
Villanova (N=34)	0.3226	0.0627+	0.1762	0.3188	0.2685	0.1246	0.1580	0.3721	0.4028	0.0182*
Wayne State (N=26)	-0.1198	0.5600	-0.2704	0.1815	-0.1619	0.4293	-0.2357	0.2464	-0.1341	0.5138

+ = significant at .10

* = significant at .05

There are 95 total correlation coefficients and only 12 are statistically significant. Each of those are positive and above .25. These positive correlation coefficients tend to cluster around four schools (St. John’s has four of them, Ohio State three, Penn State and Villanova two apiece). This suggests that at these schools the most prolific and well cited authors tended to be 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. Note also that the UCLA practice oriented number is both statistically significant at .05 and meets the .8 power requirement.

B. Pearson Correlations

The Spearman rank order correlation places each of the values in rank order and the tests whether two different sets of rank ordered values correlate. The Spearman rank order correlation does not take the gaps between the values into account, so it is considered a less powerful form of correlation coefficient than the Pearson correlation, which considers the relative values.

The Pearson correlation calculation, however, is parametric, which means that both sets of values considered must fit the assumption of normality. In order to use non-normal data those data must be transformed into normal data. Much of the data considered here is non-normal, so I transformed the data I could in order to use the Pearson correlation. If I was unable to find a transformation to normalize the data I note it in the text, and then generally used the transformation that came closest to creating normal data.

I adjusted the data using the following procedure. First I tested all data for normality using the Shapiro-Wilks test.¹⁸ Any data that was non-normal I attempted to transform. Wherever possible I used what Stata calls a lognormal zero skew transformation.¹⁹

The charts below show what transformations were used, and notes if the data could not be transformed to normality. The mass Pearson correlations showed the following:

¹⁸ I did this with the “swilk” command in Stata.

¹⁹ The Stata command for this transformation is “lnskew0”. The lognormal zero skew transformation creates a new variable that equals $\log(+/-\exp - k)$, choosing k and the sign of \exp so that the skewness of the new variable is zero.

	<u>Index Pearson</u> ²⁰	<u>P-Value</u>	<u>Zscore Pearson</u> ²¹	<u>P-Value</u>
Scholarly Productivity: ²²	0.0502	0.2112	0.0942	0.0187
Practice-Oriented:	0.0648	0.1060	0.1087	0.0066
Raw Publications Count:	0.0525	0.1903	0.0918	0.0220
Total Citations:	0.0070	0.8618	0.0427	0.2869
Citations per Year:	0.0446	0.2664	0.0802	0.0453

N= 623

The Index correlation coefficients are all so close to zero that they are a clear finding of no correlation. The sample size is large enough to find a correlation as small as .1 at a power of .8. The Z-scores, however, show a statistically significant and slightly positive correlation across four of the five measures. This is consistent with the earlier Spearman correlations. The transformed z-scores are still not normal, so a reader may want to consider those coefficients accordingly. The coefficients are basically consistent with the earlier Spearman coefficients, however, so there is an argument to be made for a small positive correlation between scholarship produced and total citations and the student evaluation z-scores. Interestingly, the small correlation between practice-oriented scholarship and teaching evaluation Z-scores is the most pronounced across both sets of correlations, and citation totals is the least likely measure of scholarship to correlate to teaching evaluations.

The school-by-school breakdown of the adjusted data shows a similar pattern. Most of the evaluation data were normal, unless marked. As noted above, I tried to transform the data using the Stata lognormal zero skew transformation wherever possible, so if the various scholarship data are not separately marked the data was transformed to normality with the lognormal zero skew transformation. The correlation coefficients that are significant at .1 are shown by a p-value in bold and italics:

TABLE 3 – TRANSFORMED PEARSON	Scholarly	P-Value	Practice	P-Value	Total Pubs	P-Value	Cites Total	P-Value	Cites Per Year	P-Value
Colorado (N=30)	0.1883	<i>0.3190</i>	0.2073	<i>0.2718</i>	0.1829	<i>0.3333</i>	0.2120	<i>0.2608</i>	0.2559	<i>0.1723</i>
Connecticut (N=32)	0.0409	<i>0.8240</i>	0.0313	<i>0.8651</i>	0.0138+	<i>0.9401</i>	0.1407	<i>0.4423</i>	0.1304	<i>0.4769</i>
Cumberland (N=25)	0.2696	<i>0.1925</i>	0.1898	<i>0.3635</i>	0.2256	<i>0.2783</i>	0.0659	<i>0.7544</i>	0.0013	<i>0.9952</i>
Florida (N=39)	-0.0524	<i>0.7515</i>	-0.1091	<i>0.5087</i>	-0.0864	<i>0.6010</i>	-0.0321	<i>0.8461</i>	0.0907	<i>0.5829</i>
Iowa (N=38)	-0.1643	<i>0.3244</i>	0.0027	<i>0.9871</i>	-0.1189	<i>0.4769</i>	-0.0240	<i>0.8864</i>	0.0071	<i>0.9663</i>
Lewis & Clark (N=29)	0.0242	<i>0.9009</i>	0.0919	<i>0.6352</i>	0.1519	<i>0.4315</i>	0.2067	<i>0.2820</i>	0.1807	<i>0.3483</i>
Michigan (N=35)	0.2194	<i>0.2055</i>	0.1856	<i>0.2858</i>	0.1985	<i>0.2530</i>	0.1889	<i>0.2771</i>	0.1894	<i>0.2759</i>
UND (N=10)	- 0.1498*	<i>0.6796</i>	0.0024*	<i>0.9947</i>	-0.047*	<i>0.8965</i>	0.3319+	<i>0.3487</i>	0.3956	<i>0.2579</i>

²⁰ The Index data were not normal and were adjusted using a lognormal zero skew transformation.

²¹ I could not normalize the zscores. A zero skew lognormal transformation most closely approximated a normal distribution and was used for these correlations.

²² None of the data for Scholarly Productivity, Practice Oriented, Raw Publications, Total Citations, or Citations per Year were normal. A log normal zero skew transformation was used, although the transformed Scholarly Productivity, Practice Oriented, Raw Publications, and Citations per Year remain non-normal even after the transformation. The transformed Total Citations data are normally distributed.

Northwestern ²³ (N=41)	0.0036	0.9822	-0.0222	0.8904	-0.0030	0.9849	-0.1868	0.2422	-0.0987	0.5394
Ohio State (N=35)	0.3204	0.0606	0.3562	0.0357	0.2982#	0.0819	0.0970	0.5795	0.1511	0.3863
Penn State ²⁴ (N=24)	0.4396	0.0316	0.3902	0.0594	0.3048	0.1476	0.1142	0.5952	0.2526	0.2337
Southwestern (N=38)	-0.1650#	0.3223	-0.1502#	0.3680	-0.1967#	0.2366	-0.1764	0.2894	-0.0033	0.9843
St. John's ²⁵ (N=37)	0.3361	0.0420	0.2790	0.0944	0.3176	0.0554	0.1476	0.3833	0.2957	0.0756
Tennessee ²⁶ (N=35)	-0.0048	0.9780	0.3168	0.0637	0.1136	0.5159	0.1156	0.5083	0.0670	0.7022
Texas Tech (N=33)	-0.1060#	0.5570	-0.0992	0.5827	-0.0767	0.6715	-0.0119	0.9476	-0.1322	0.4635
Toledo (N=24)	-0.1099#	0.6092	-0.1777#	0.4062	-0.1201#	0.5763	0.2121	0.3197	0.1999	0.3491
UCLA ²⁷ (N=58)	0.1668	0.2108	0.3430	0.0084	0.1680	0.2076	-0.0059	0.9651	-0.0677	0.6134
Villanova (N=34)	0.3533	0.0404	0.1845	0.2962	0.2963	0.0889	0.1946	0.2701	0.4154#	0.0146
Wayne State (N=26)	-0.0993+	0.6293	-0.3559	0.0743	-0.1710+	0.4035	-0.2321	0.2539	-0.1751	0.3923

+ = These data were normal, and were not transformed
 * = log (x +1) transformation was used to achieve normality²⁸
 # = Log normal zero-skew transformation used, but data remained non-normal

There are 14 correlations out of 95 that are statistically significant at a .1 p-value. Again the statistically significant correlations clustered around Ohio State, St. John's and Villanova with three each, and Penn State with two. Wayne State has our first statistically significant negative value. The UCLA practice-oriented coefficient is statistically significant at .05 and meets a .8 power requirement. Again, there are no statistically significant correlations associated with total citations.

Nevertheless, despite a finding of correlation at four of nineteen schools the over-all picture supports the amalgamated numbers: there is either no correlation between teaching evaluations and these measures of scholarly output, or a very slight positive correlation. This is especially so in light of the findings of the large scale correlations and the general uniformity of the school level correlations. The Pearson correlations of adjusted data support a finding of no correlation or slight positive correlation.

Lastly, I include the unadjusted Pearson correlations. Most of these data are not normal, so these correlation coefficients should be used appropriately. Nevertheless, I include them to show their overall consistency with the earlier coefficients.

²³ Northwestern's evaluation data were not normally distributed. None of the transformations resulted in normal distributions. The log normal zero skewness transformation most closely approximated a normal distribution, so I used those data.

²⁴ Penn State's evaluation data were not normally distributed, so a Box-Cox transformation was used. The Box Cox transformation creates a new variable that equals $(\exp^a - 1)/a$. The Box-Cox transformation was not used for other data, because exp must be strictly positive to use the Box-Cox transformation and the other data either contained zeros or negative values. The Stata command for a Box-Cox zero skew transformation is `bcskew0`.

²⁵ St. Johns' evaluation data were not normally distributed, so a Box-Cox transformation was used.

²⁶ Tennessee's evaluation data were not normally distributed, so a Box-Cox transformation was used.

²⁷ UCLA's evaluation data were not normally distributed, so a Box-Cox transformation was used.

²⁸ This means I calculated a new variable that equals $\log(\exp+1)$.

Here are the large-scale, unadjusted correlations for the index and z-scores, as well as the unadjusted Pearson coefficients for each school:

	<u>Index Pearson</u>	<u>P-Value</u>	<u>Z-score Pearson</u>	<u>P-Value</u>
Scholarly Productivity:	.0029	.9417	.0383	.3403
Practice-Oriented:	.0366	.3614	.0650	.1053
Raw Publications Count:	.0229	.5677	.0528	.1885
Total Citations:	-.0139	.7298	.0194	.6293
Citations per Year:	.0224	.5770	.0487	.2251

N= 623

TABLE 4 – Unadjusted Pearson

	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

Out of the 95 correlations there are only 8 that are significant at .10. Similarly the amalgamated numbers show correlation coefficients very close to zero and generally high p values indicating that even the small correlation coefficients found are not statistically significant.

In addition to the mass and individual correlations, I also created scatterplots for each data set, and reviewed them in case there was an undetected curvilinear relationship not revealed by the linear correlations. I have posted these scatterplots on a specified website for review.²⁹ None of the scatterplots show a strong curvilinear relationship.

²⁹ See <http://www.law.utk.edu/FACULTY/BartonScatter.pdf>.

Because the earlier work of Lindgren & Nagelberg found a small relationship between citations and teaching evaluations for the most and least cited professors at four schools, and also that the more cited professors were more likely to get good teaching evaluations, I considered whether the top 20% and the bottom 20% of the total citations count had a stronger or weaker relationship than the entire cohort. The results showed, if anything, less of a relationship:

BOTTOM 20% total cites:

	<u>Index Spearman</u>	<u>P-Value</u>	<u>Z score Spearman</u>	<u>P-Value</u>
Scholarly Productivity:	0.0163	0.8564	0.0591	0.5124
Practice-Oriented:	0.0002	0.9979	0.0401	0.6567
Raw Publications Count:	0.0180	0.8419	0.0475	0.5986
Total Citations:	0.0877	0.3310	0.1058	0.2404
Citations per Year:	0.0620	0.4920	0.0959	0.2873

N= 125

TOP 20% total cites:

	<u>Index Spearman</u>	<u>P-Value</u>	<u>Z score Spearman</u>	<u>P-Value</u>
Scholarly Productivity:	0.0383	0.6713	0.0513	0.5701
Practice-Oriented:	0.1209	0.1792	0.1157	0.1988
Raw Publications Count:	0.0283	0.7543	0.0140	0.8767
Total Citations:	-0.0796	0.3776	-0.0643	0.4762
Citations per Year:	0.0316	0.7263	0.0020	0.9820

N= 125³⁰

In sum, across the different calculations of scholarly productivity and citation, and using both a linear index score and a z-score to combine the teaching evaluation data the findings show either no correlation or a slight positive correlation. In light of the slightness of the few correlations that were found the safest read across this data is that teaching evaluations do not correlate strongly with either scholarly productivity (whether measured as a raw publication count or weighted) or a citation count.

It is interesting to note that this finding of no or slight correlation is basically repeated for most schools on the various school-by-school charts above. Except for a few schools (which are dissimilar enough from each other that they cannot be easily grouped or

³⁰ I also considered the effect of outliers on these findings. The use of the Spearman correlation coefficient and the use of the transformed data for the Pearson correlations, however, make the existence of outliers much less worrisome. This is because the Spearman correlation places each data point in rank order rather than considering the absolute value, and using the transformed data tends to bring the outliers closer to the pack for the Pearson correlations. Nevertheless, I did run some of the correlations without an outlier or two and found little change in the results.

generalized) 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 or a weak positive correlation is also consistent with the non-law studies that made similar findings.

B. The Counter-Intuitive Nature of this Finding

In describing this study to others, I found two general predictions as to my results. A slight majority predicted that scholarly productivity and teaching effectiveness have a strong positive correlation. 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 strong 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.

Unless a person had read the literature outside of law schools that shows a general finding of no correlation in other disciplines, few 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 or very slight correlation makes it unlikely that teaching and scholarship are having a large effect on 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 who 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 who 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, at least these data suggest that it seems to have little effect, or a slightly positive 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 (ABA Standards 2006; AALS By-Laws 2006), if scholarship and teaching effectiveness are not strongly related. 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.³¹

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.

³¹ 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.

REFERENCES

AALS, AALS By-Law Section 6-4(a).

AALS, Member Schools, http://www.aals.org/about_memberschools.php (last visited June 20, 2006).

ABA, ABA Standards for Approval of Law Schools, Standard 404.

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).

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

Braxton, John M. 1996. Contrasting Perspectives on the Relationship Between Teaching and Research. Pp. 5-14 in *Faculty Teaching and Research: Is There a Conflict*, edited by John M. Braxton. San Francisco: Jossey-Bass Publishers.

Cope, David. 2005. *Fundamentals of Statistical Analysis*. New York, NY: Foundation Press.

Cullen, Colleen M. & Randall S. Kalberg. 1995. Chicago-Kent Law Review Faculty Scholarship Survey, *Chicago-Kent Law Review*, 70: 1445-98.

Dickson, Jean D. 1993. *Nonparametric Measures of Association*. London, UK: Sage Publications.

Eisenberg, Theodore & Martin T. Wells. 1998. Ranking and Explaining the Scholarly Impact of Law Schools. *Journal of Legal Studies*, 27: 373-88.

Eisenberg, Theodore & Martin T. Wells. 2000. Inbreeding in Law School Hiring: Assessing the Performance of Faculty Hired From Within. 29 *Journal of Legal Studies*, 29: 369-99 (2000).

Farley, Christene H. 1996. Confronting Expectations: Women in the Legal Academy. *Yale Journal of Law and Feminism*, 8: 333-73.

Feldman, Kenneth A. 1987. Research productivity and scholarly accomplishment of college teachers as related to their instructional effectiveness: A review and exploration. *Research in Higher Education*, 26: 227-298.

Hattie, John A.C. & Herbert W. Marsh. 1996. The Relationship Between Research and Teaching—A Meta-Analysis. *Review of Educational Research*, 66: 507-542.

Hattie, John A.C. & Herbert W. Marsh. 2002. The Relation Between Research Productivity and Teaching Effectiveness—Complementary, Antagonistic, or Independent Constructs? *Journal of Higher Education*, 73: 603-641.

Korobkin, Russell. 1998. In Praise of Law School Rankings: Solutions to Coordination and Collective Action Problems. *Texas Law Review*, 77: 403-28.

Leiter, Brian. 2000. Measuring the Academic Distinction of Law Faculties, *Journal of Legal Studies*, 29: 451-98.

Leiter, Brian. Faculty Quality Based on Scholarly Impact, http://www.leiterrankings.com/faculty/2005faculty_impact_cites.shtml (last visited March 30, 2006).

Leonard, James. 1990. Seein' the Cites: A Guided Tour of Citation Patterns in Recent American Law Review Articles. *Saint Louis Urban Law Journal* 34: 181-220.

Lindgren, James & Nagelberg, Allison. 1998. Are Scholars Better Teachers? *Chicago-Kent Law Review*, 73: 823-33.

Marsh, H. W. (2007). Students' evaluations of university teaching: A multidimensional perspective. In R. P. Perry & J C. Smart (Ed.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective* (pp. 319-384). New York: Springer.

Marsh, H.W. (1984). Students' evaluations of teaching: Dimensionality, reliability, validity, potential biases and utility. *Journal of Educational Psychology*, 76, 707-754.

Marsh, H.W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for further research. *International Journal of Educational Research*, 11, 253-388.

Marsh, H.W., & Overall, J.W. (1979). *Validity of students' evaluations of teaching: A comparison with instructor self evaluations by teaching assistants, undergraduate faculty, and graduate faculty*. Paper presented at the annual meeting of the American Educational Research Association. (ERIC Document Reproduction Service, No. ED 177 205).

Merritt, Deborah J. 1998. Research and Teaching on Law Faculties: An Empirical Exploration. *Chicago-Kent Law Review*, 73: 765-821.

Newton, Rae R. & Rudestam, Kjell E. 1999. *Your Statistical Consultant*. London, UK: Sage Publications.

O'Reilly, Mark T. 1987. Relationship of Physical Attractiveness to Students' Ratings of Teaching Effectiveness. *Journal of Dental Education*, 51: 600-602.

Scordato, Marin R. 1990. The Dualist Model of Legal Teaching and Scholarship. *American University Law Review*. 40: 367-95.

SISA, Sample Size Calculation, <http://home.clara.net/sisa/correl.htm> (last visited July 27, 2007).

Slomanson, William R. 2000. Legal Scholarship Blueprint, *Journal of Legal Education*, 50: 431-50.

Smith, Pamela J. 1999. Teaching the Retrenchment Generation: When Sapphire Meets Socrates at the Intersection of Race, Gender, and Authority. *William and Mary Journal of Women and the Law*. 6: 53-195.

The University of Tennessee, TN101 Online, <http://tn101.utk.edu/default.asp> (last visited March 30, 2007).