

University of Chicago Law School

Chicago Unbound

Coase-Sandor Working Paper Series in Law and Economics Coase-Sandor Institute for Law and Economics

2019

What Should Law School Rankings Measure and How Should We Measure it: A Comment on Heald & Sichelman's Rankings

Adam Chilton

Jonathan S. Masur

Follow this and additional works at: https://chicagounbound.uchicago.edu/law_and_economics



Part of the **Law Commons**

**WHAT SHOULD
LAW SCHOOL RANKINGS MEASURE
AND HOW SHOULD WE MEASURE IT:
A COMMENT ON
HEALD AND SICHELMAN'S RANKINGS**

Adam Chilton* and Jonathan Masur**

ABSTRACT: There are obvious benefits to ranking academic departments based on objective measures of faculty research output. However, there are considerable difficulties associated with producing reliable and accurate rankings. In this short comment, we offer an evaluation of Heald and Sichelman's recent foray into the project of ranking law schools. Heald and Sichelman are to be commended for the transparency and rigor of their rankings effort. At the same time, it is important to note that their rankings involve a series of contestable discretionary choices and could give rise to potential counterproductive gaming by law schools seeking to improve their place in the rankings. In particular, Heald and Sichelman's system places a thumb on the scale on behalf of more senior faculty who publish in traditional law reviews and write in popular substantive areas like constitutional law. This raises the concern that rankings of this type could discourage law schools from hiring faculty that are young, produce interdisciplinary scholarship, and write in otherwise underrepresented fields. Nonetheless, Heald and Sichelman have taken an important step forward, and other scholars should look to build profitably upon their work.

CITATION: Adam Chilton & Jonathan Masur, *What Should Law School Rankings Measure and How Should We Measure It: A Comment on Heald and Sichelman's Rankings*, 60 JURIMETRICS J. __–__ (2019).

There are obvious benefits to ranking academic departments based on objective measures of faculty research output. However, there are considerable difficulties associated with producing reliable and accurate rankings. Paul Heald and Ted Sichelman deserve substantial credit for undertaking a project to improve upon existing efforts to rank the academic impact of law school faculties, particularly in light of the fact that *U.S. News & World Report* now claims that it will incorporate some type of citation ranking into its overall assessment of law schools.¹ Moreover, Heald and Sichelman's work is careful and thorough, and they are admirably honest about its limitations and shortcomings. They also deserve considerable credit for their commitment to transparency and replicability.

*Professor of Law, Walter Mander Research Scholar, University of Chicago Law School.

**John P. Wilson Professor of Law, David and Celia Hilliard Research Scholar, Director of the Wachtell, Lipton, Rosen & Katz Program in Behavioral Law, Finance and Economics, University of Chicago Law School.

1. Paul J. Heald & Ted Sichelman, *Ranking the Academic Impact of 100 American Law Schools*, 60 JURIMETRICS J. (forthcoming 2019).

At the same time, there are many issues raised by their rankings that deserve further discussion. The reason is simple: if rankings of academic impact based on citations are incorporated into *U.S. News*'s overall rankings of law schools, schools may try to maximize their ranking, even at the cost of other academic or social values. Decisions regarding how to construct rankings will thus directly impact the kinds of scholarship, and scholars, that are produced and rewarded. These decisions should thus be carefully debated by the legal academy and only made after thoughtful consideration about exactly what the rankings should measure and how best to do it. To contribute to that necessary debate, in this short comment, we have two goals: to describe and decompose some of those necessary decisions about what to measure, and to highlight some problems that may result from using HeinOnline to measure faculty impact.²

I. DECIDING WHAT TO MEASURE

It is impossible to develop law school rankings without making normative judgements about what should be measured. To their credit, Heald and Sichelman do not ignore this reality. Instead, they argue that rankings can be used for several purposes: (1) provide a general indication of faculty reputation; (2) aid entry-level and lateral job candidates who are making decisions about particular law schools; and (3) provide advice to students “who care about the scholarly reputation and quality of their professors.”³ Identifying these purposes raises many questions about what rankings should measure, but it does not provide clear answers to them. Their current rankings make at least four contestable judgements that should potentially be reexamined.

First, in addition to including citations from HeinOnline, the Heald and Sichelman rankings are also based on downloads from SSRN. Although there are reasons to be concerned about the way HeinOnline measures citations (which we discuss below), the basic case for using citations to assess academic impact is straightforward: they are a revealed preference measure of the impact

2. We should note at the outset that our own scholarship is not entirely immune from our criticisms. See, e.g., Adam Chilton, Jonathan Masur, & Kyle Rozema, *Rethinking Law School Tenure Standards* (Sept. 17, 2019) [hereinafter Chilton et al., *Tenure Standards*] (unnumbered working paper), https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3200005; Adam Chilton, Jonathan Masur, & Kyle Rozema, *Affirmative Action in Law Reviews* (Dec. 30, 2018) (unnumbered working paper), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3295334; Adam Chilton, Jonathan Masur, & Kyle Rozema, *Political Discrimination in the Law Review Selection Process* (Jan. 31, 2019) (unnumbered working paper), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3119903. We would note two things in our defense against the charge of hypocrisy we have just leveled at ourselves. First, the goal of our scholarship has been to test hypotheses, and not to produce rankings. As a result, many of the limitations in our own work are a source of measurement error that should be acknowledged, but we do not make relevant claims about specific observations (e.g., School A compared to School B) because we recognize the data may be too noisy for such a specific comparison. However, highlighting these specific comparisons is the exact purpose of producing rankings. Second, we hope the debate generated by Heald and Sichelman's important project will improve many scholars' future work—including our own.

3. Heald & Sichelman, *supra* note 1 (manuscript at 37). We're not sure of the value of the first of these purposes independent of the second and third, except to provide a sense of which law schools are “winning,” so we will leave that one aside and focus on the second and third purposes.

of one piece of scholarship on other scholarship. The case for using SSRN downloads as a proxy for academic impact, however, is less straightforward. The case that Heald and Sichelman make for using SSRN downloads is that they provide a useful corollary to citations as a measure of academic impact, because

[citation databases] catch formal scholarly references to an article, but they provide no indication of how often the article has been read (rather than cited) not only by legal academics but also by practitioners and others outside the legal academy. By incorporating SSRN data, we provide a better sense of the “splash” an article has made. We argue that an article that has been downloaded 1000 times is likely to have enhanced its author’s reputation more than an article that has been cited 10 times.⁴

Although we would not disagree that 1000 downloads may do more to enhance reputation than 10 citations, it is not clear that academic impact rankings should be trying to measure reputation in this way. SSRN downloads occur before the paper is read (and thus before the quality, novelty, or relevance of the argument is assessed), and downloads are often driven by either the visibility of the author or links from the popular press. Thus, downloads may tell us little about the underlying quality of an article or its impact on academic debates. Or, to put it another way, acquiring 1000 Twitter followers may do more to enhance a law professor’s reputation than writing an article that is cited 10 times, but most would likely agree that counting law professors’ Twitter followers would be a bad way to measure their quality. Similarly, it certainly enhances a law professor’s reputation when she publishes an op-ed, is cited by courts, or testifies before congress. Like downloads, these activities may be important, but they are different than academic impact or quality as it has traditionally been measured. Moreover, given the correlation between Heald and Sichelman’s SSRN-only and Hein-only rankings (0.84),⁵ it is not clear the inclusion of SSRN is necessary.

Second, the Heald and Sichelman rankings do not make adjustments for fields, age, or other demographic factors. Although citations have not been studied as extensively in law as they have in other fields, there is evidence that citations vary by factors like age, field, gender, and race. As a result, the choice to not adjust for these features is an implicit decision about what kinds of faculty composition should be rewarded. For instance, we have found in our own work that tax scholars are cited less than constitutional law scholars.⁶ If rankings are not adjusted for field, law schools might decide to hire fewer tax scholars and more constitutional law scholars, because constitutional law scholars generally earn more citations. This trend will be self-reinforcing: as the number of constitutional law professors increases, the number of constitutional law papers will increase, thereby increasing the number of citations of work by other constitutional law professors. Tax citations will decrease via the opposite mechanisms. This will skew law schools’ incentives even further.

4. *Id.* (manuscript at 4–5).

5. *Id.* (manuscript at 5).

6. Chilton et al., *Tenure Standards*, *supra* note 2, at 34.

In addition, rankings that rely on the median or mean of cumulative lifetime citations will predictably quite heavily reward schools with older faculties. The question of whether it makes sense to favor older faculties in this manner depends greatly upon the purposes the rankings are meant to serve. Of course, some students and faculty candidates might be seeking the faculties with the greatest lifetime accomplishments. If this is the case, then cumulative lifetime citations is the most sensible measure. But other students and faculty candidates might be more interested in the faculties currently producing the best work. If this is the case, citation counts from the past one to five years would be a superior measure.

However, even rankings that look only at citations within the past one to five years will favor older faculties, possibly by significant margins. The reason is that those faculties will have a larger accumulated body of work that might be cited. A faculty member who has been teaching for 20 years and has published 40 articles is likely to garner more citations within the next twelve months than a faculty member who has been teaching for 5 years and has published 10 articles, all else being equal. Heald and Sichelman's data supports this intuition.⁷

There are several approaches that could be used to counteract the thumb on the scale favoring older faculties and provide a more accurate picture of which law schools are currently the most productive. One would be effectively to control for seniority by grouping scholars into age-related cohorts and evaluating citations within cohorts. That is, a ranking system could compare citations to scholars who have been teaching zero to five years, six to ten years, eleven to fifteen years, and so forth. Cohorts could then be compared across schools to evaluate how the most junior scholars at School X performed compared to the most junior scholars at School Y, how mid-career scholars at the two schools performed, and so on.⁸ Another approach would be to count citations within the past *X* years *only to articles published within the past X years*. This would provide an indication of the quality of work being done by a faculty—including both senior and junior scholars—at the current moment, rather than a ranking that largely reflects work done years or decades earlier.⁹

Third, the Heald and Sichelman rankings attribute citations entirely to scholars and not to the place where the scholarship was produced. For example, imagine a highly influential faculty member, Professor Smith, who has taught at School B since 2015. Before 2015, she taught at School A. Professor Smith was already a highly influential scholar with many citations when she moved

7. See Heald & Sichelman, *supra* note 1 (manuscript at 15). Their finding that twelve-month SSRN download numbers do not favor older faculties is unsurprising. There is much less reason to download an article on SSRN once that article has been published because it is available through other means. Accordingly, older articles will generally reap many fewer downloads than newer articles, and a senior scholar's larger corpus of existing scholarship will not provide as much of an advantage.

8. For example, we used this kind of cohort-based grouping in our study on tenure standards to ensure that professors were being compared to their contemporaries. See Chilton et al., *Tenure Standards*, *supra* note 2, at 3–5.

9. We also used a variant of this methodology in our tenure study, where we examined pre-tenure citations to pre-tenure articles and post-tenure citations to post-tenure articles. *Id.*

from School A to School B in 2015, and the work she did at School A—including a particularly well-cited 2011 article, *A Theory of Law*—is still cited quite regularly. The Heald and Sichelman rankings would attribute all of Professor Smith's citations to School B, because that is where she teaches now. In particular, if *A Theory of Law* was cited 100 times in 2013 (while Professor Smith was teaching at School A), Professor Smith would “carry” those citations with her when she moved to School B, even though she earned them while she was teaching at School A based on work done at School A.

Whether this decision is reasonable depends on what the rankings are trying to measure. For instance, faculty and students might be seeking very different things when they choose a law school—and, of course, different faculty and different students might be seeking different things from one another as well. Some faculty or students might be seeking schools with the most prestigious and well-regarded senior faculty, perhaps on the theory that they have the greatest body of accumulated knowledge. Other faculty or students might instead be looking for the faculty that is currently most active in scholarship and producing the most (and best) new work, perhaps on the theory that such an environment will be most conducive to research and learning. The design of the rankings will influence which of these priorities is better captured. In the example above, it is the students and faculty of School B who now have the privilege of interacting with Professor Smith, and so perhaps all of her influence and accumulated knowledge—as proxied for by those citations—should be attributed to School B. On the other hand, School A obviously did something right before 2015 when it hired Professor Smith and created the conditions that allowed her to write such an important piece of scholarship. A new faculty member seeking a generative environment might prefer a system of rankings that attributed Professor Smith's work at School A to School A, thus reflecting the productivity of scholars who are members of that faculty.

Additionally, attributing citations to the place they were “earned” (in whole or in part) might mitigate perverse hiring and retention incentives. Specifically, if citations are attributed entirely to the scholar, schools could improve their ranking by hiring a senior lateral candidate who is no longer research active instead of hiring a more junior lateral candidate who is more engaged in scholarship and teaching. This type of move might be good for a law school's ranking but bad for the students and other faculty at the law school. Relatedly, schools will have good reason to encourage senior faculty members to not retire—even if they are no longer effective teachers, mentors, or scholars—simply to retain their citations for *U.S. News* rankings. Both behaviors would limit the opportunities for new scholars to break into the legal academy and, in turn, hamper the production and dissemination of knowledge. It would also further entrench the status quo distribution of academic jobs along existing race and gender lines.

Fourth, there is the further question, which Heald and Sichelman explore, regarding whether it makes more sense to use the mean or median citations of a faculty as a measure of that faculty's quality. The use of mean citations will favor faculties with right-tail outliers—true superstars—even if the remainder

of the faculty is weaker. The use of median citations will favor faculties of more consistent overall quality.

Heald and Sichelman have taken what largely amounts to a middle route through these questions. As their measure of citations, they use $2 \times \text{mean} + \text{median}$. They also give equal weight to total lifetime citations and citations within the past twelve months. However, both of those measures favor more senior faculties. This is of course just a first effort, and a single paper, and thus Heald and Sichelman should not be criticized for their attempts to balance these competing considerations. To the contrary, they deserve praise for their commitment to making the raw data available so future researchers can try alternative weighting schemes.

Nonetheless, it might be that this middle-ground approach, in attempting to serve all possible purposes, does not serve any of them especially well. It might be more useful to have two (or more) separate rankings, each tailored to various purposes. We could imagine a “lifetime” ranking that looked at mean lifetime citations; a “current productivity” ranking that looked only at contemporary citations to contemporary articles; and possibly a balanced ranking that controlled for seniority by grouping faculty into age cohorts. Again, the choice between these options depends very much upon what purpose the rankings are meant to serve.

II. ACCURATE MEASUREMENT

Whatever rankings are trying to measure, it matters that they measure it accurately. This is because we should anticipate that schools will likely try to game rankings, just as schools have tried to game all other aspects of *U.S. News* rankings. It would be especially pernicious if rankings created incentives for law schools to exploit measurement error. Thus, it is important to try to insulate the rankings from being gamed as much as possible, or at minimum to try and ensure that “gaming” the rankings will result in schools maximizing the types of things our profession deems worth maximizing.

Here, the most important limitation of HeinOnline—one that Heald and Sichelman discuss—is that it includes only citations to articles within the Hein database by articles within the Hein database. That means that these rankings will dramatically underestimate the influence of scholars who regularly publish or are cited in other types of publications, such as books or nonlegal journals. Both of these limitations—related to the venue of publication, and the source of the citation—will tend to disadvantage scholars who work at the intersection of law and cognate disciplines, particularly history and philosophy (where much of the scholarly output occurs through books), but also economics, political science, psychology, sociology, and others.

Heald and Sichelman admirably acknowledge the limitation related to scholarly work that is published in books and (even more admirably) resolve to correct it in future versions of their study. However, they are somewhat dismissive of the value of citations to legal work in nonlegal publications.¹⁰ We are

10. Heald & Sichelman, *supra* note 1 (manuscript at 10) (“Even so, it is not clear to us that out-of-field citations—for instance, in the anthropology or psychology literature—have the same relevance for measuring law school impact as in-field citations.”).

not sure why a citation to a law professor's work in a philosophy or history journal should count less than a citation to the same individual's work in a law journal. The point of our profession is to create and disseminate knowledge, and knowledge in a cognate discipline seems no less valuable than knowledge in law itself. This is especially true in an era in which more and more law faculty have training in related fields and do work that spans both fields. To take just one somewhat extreme example: Richard Thaler's work on "nudging," which he co-authored with Cass Sunstein, was cited in the announcement of his Nobel Prize in Economics.¹¹ The Royal Swedish Academy of Sciences obviously did not consider Thaler's work to be less valuable merely because a significant portion of its impact has been felt in law, rather than economics. It would be equally bizarre to discount the impact of Cass Sunstein's contributions to that same body of scholarship because some of that impact has been felt in economics, rather than law.

Any source of measurement error will likely replicate itself in the real world. If we produce and rely upon rankings that do not accurately measure the impact of historians, this will have the recursive effect of (incorrectly) reducing the number of legal historians hired by law faculties. It is thus important to find solutions to these problems. The most natural solution would be to require every school to produce Google Scholar pages for every scholar, as Brian Galle has suggested.¹² But other solutions, such as the full-text analysis Heald and Sichelman mention, should be explored as well.



We believe that the questions we have raised here are fundamental to any attempt to rank law faculties. But at the same time, we do not believe that there are objectively right answers to most of these questions. As we have said, the right answer might depend upon who is using the rankings, and for what purpose. As things stand, the important normative judgments are being made unilaterally by the rankers themselves—Sisk, Heald and Sichelman, and (soon) *U.S. News*. This is not necessarily undesirable, but it might be preferable if they were made through a sort of deliberative process by our discipline as a whole. It would be even better if they could be made behind the veil of ignorance, with law professors deciding on which measures are most sensible without knowing how they (or their schools) would fare under those measures. That is likely impossible, but one can always dream. In the meantime, Heald and Sichelman are to be commended for an important and significant step in the right direction.

11. Press Release, Royal Swedish Acad. of Scis., The Prize in Economic Sciences 2017 (Oct. 9, 2017), <https://www.nobelprize.org/uploads/2018/06/press-43.pdf> [<https://perma.cc/G48Y-2799>].

12. Brian Galle, *How I'd Fix the Sisk-Leiter Citation Studies*, MEDIUM (May 25, 2016), <https://medium.com/whatever-source-derived/how-id-fix-the-sisk-leiter-citation-studies-4489f87ed387#n60eh4jf6> [<https://perma.cc/K4CR-NXJ6>].