

Are great teachers poor scholars?

David N. Figlio and Morton Schapiro

Executive Summary

Colleges and universities must balance many goals, and research universities in particular aspire to excellence in both teaching and research. University administrators and policymakers alike are interested in ensuring that publicly-supported private and public universities operate at high levels of instructional and scholarly quality, but to date we know little about whether scholarly excellence comes at a cost in terms of teaching quality, or vice versa.

We bring to bear unique matched student-faculty data from Northwestern University, a mid-sized research university that is one of the 26 private universities among the 62 members of the Association of American Universities, to investigate the relationship between teaching and scholarly quality. Using the full population of all first-year undergraduates enrolled at Northwestern between fall 2001 and fall 2008 (over 15,000 students in all), we empirically generate two new measures of teaching quality—one an indicator of *inspiration* (the rate of “conversion” of non-majors to majors) and the other an indicator of *deep learning* (the degree to which a professor adds lasting value to students’ learning that is reflected in success in future classes). We also investigate two measures of research quality—one based on a measure of the relative importance of a scholar’s research in the field, and the other a measure of national or international prominence as reflected by major awards.

We find that, regardless of our measure of teaching quality or our measure of research quality employed, there is no relationship between the teaching quality and research quality of tenured Northwestern faculty. Our estimates are “precise zeroes,” indicating that it’s unlikely that mismeasurement of teaching or research quality explains the lack of a relationship between the two. Therefore, while Northwestern admittedly occupies a rarefied space in the hierarchy of American universities, our results suggest that excellent teaching and excellent research are not substitutes (though neither are they apparently complements).

Colleges and universities have a variety of output goals. At a research university, scholarly output is critical, but so is successful teaching at the undergraduate, professional school, and graduate levels. Some faculty members are fabulous at both core tasks. They are revered by undergraduates for their ability in the classroom, even when teaching at the most basic, introductory level. At the same time, they are widely recognized by their peers for their research excellence. On the other hand, some faculty are much more proficient in publishing than in teaching, while some excel mainly in the classroom. Others, unfortunately, struggle at both.

Given how important these different desired outcomes are for university administrators and policymakers interested in ensuring that their publicly-supported private and public universities operate at high levels of instructional and scholarly quality, one would think that deans, provosts, and presidents would know quite a bit about the production of both top-notch research and top-notch teaching, and, in particular, they would know whether faculty who are superstars in the undergraduate classroom pay a price in terms of scholarly recognition.ⁱ Knowing the answer to this question could help institutions better allocate their scarce resources.

Unfortunately, that kind of analysis is hampered by two issues—the difficulty in measuring teaching effectiveness, and the difficulty in comparing research success across a wide range of academic fields. A long literatureⁱⁱ on the relationship between teaching and research excellence spanning four decades (including papers published recentlyⁱⁱⁱ) considers various measures of each of these professorial attributes. Previous studies measure research excellence based on number of publications, grants awarded, number of citations, peer/department chair rating, time spent on research activities, faculty membership in academic research societies, and awards for research. While some of these measures are plausible indicators of research quality, others are either quite narrow or do not promote comparability across disciplines. Meanwhile, studies in the existing literature have been even more hampered by limitations in measuring teaching quality. Previous analyses of the teaching-research relationship have used student course evaluations, peer evaluations, time spent on teaching activities, and nomination or receipt of a teaching award as teaching quality measures,^{iv} but these measures have substantial limitations. Student evaluations frequently reflect popularity or grading standards, rather than genuine instructional quality, and exhibit gender, racial, and ethnic biases,^v while

peer reviews are subject to “halo effects” resulting from evaluators’ knowledge of a faculty member’s research record.^{vi} What is needed are measures of teaching effectiveness that are not subject to these types of biases.

In this report we propose two measures of teaching outcomes and two indicators of research prominence in order to directly assess the relationship between teaching and research excellence. Our measures of research prominence are along the lines as those employed in previous work, but with some refinements to allow us to compare across disciplines. We empirically derive our measures of teaching excellence from data on student follow-on course-taking and future performance. One of the measures of teaching quality is an indicator of *inspiration* while the other is an indicator of *deep learning*.

We study the teaching and research quality of tenured faculty members at Northwestern University, a midsized research university that is one of the 26 private universities among the 62 members of the Association of American Universities. We construct our teaching measures using registrar data on all Northwestern University freshmen who entered between fall 2001 and fall 2008, a total of 15,662 students, and on the faculty who taught them during their first quarter at Northwestern. In a recent paper, we document that student course-taking during that term appeared to be essentially random with regard to measures of faculty quality, providing indicators of teaching quality that are unaffected by student non-random selection.^{vii}

Methods

We focus here just on the 170 tenured faculty members who taught at least 20 first quarter students across the eight cohorts. We limit our analyses to those with 20 or more observations of first quarter students in order to reduce the likelihood that outliers are influencing our measures of teaching quality.^{viii} We concentrate on tenured faculty members because they have had longer to establish the teaching and research track records necessary to carry out this analysis, and because they have already met the high bar of teaching, research, and service necessary for tenure at a leading American research university.

We describe the first of our teaching measures in some detail in our paper cited above.^{ix} We build on the work of other researchers^x in examining the likelihood of taking additional courses in a subject area as a

measure of teaching excellence—the more likely a student is to continue studying in a particular discipline, the better the teacher is thought to have been. We then take this approach further by developing our own objective measure—the deviation in the grade received by a student in follow-up courses in that subject. A successful undergraduate teacher in, say, introductory biology, not only induces his or her students to take additional biology courses, but leads those students to do unexpectedly well in those additional classes (based on what we would have predicted based on their standardized test scores, other grades, grading standards in that field, etc.) In our earlier paper, we lay out the statistical techniques^{xi} employed in controlling for course and student impacts other than those linked directly to the teaching effectiveness of the original professor. Similar approaches to measuring instructor quality using follow-on course performance have been employed by other scholars.^{xii}

This is an imperfect measure of teaching success. An especially poor teacher may lead students to switch from that subject entirely—hence we would have no information on how their students would have done in subsequent courses. In addition, our data set centers on success in teaching courses open to first-quarter freshmen, most of which are introductory courses. In around one-fifth of cases, they are intermediate or upper-division classes taken by students with prior experience in the subject such as that gained through AP classes in high school.^{xiii} Finally, and most importantly, perhaps some faculty look to be poor teachers by our measure, but do a better teaching senior seminars for undergraduates or graduate courses. Teaching is certainly not limited to what takes place in a classroom full of first year undergraduates.

Our second indicator of teaching effectiveness uses the same data set but focuses on majors rather than grades. Compelling teachers presumably inspire students into majoring in their disciplines, whether those students were predisposed to doing so or not. A talented chemistry teacher in an introductory course may lead a declared chemistry major to keep that major; but even more impressively, that professor may convince an undeclared student to major in chemistry, or lead a student majoring in physics or economics to become a chemistry major. The ability to convert majors presumably reflects an important dimension of teaching excellence.

In essence, therefore, our two measures of teaching quality reflect, in the first case, value added (or “deep learning”) that is transferrable to subsequent classes in the subject, and, in the second case, inspiration, as

indicated by the ability to convert students to a subject that they had not previously planned on studying in depth. Interestingly, the faculty members who are most successful in inspiring students to become majors in their subject are not any more distinguished in facilitating “deep learning”: the correlation between the two measures of teaching quality is virtually zero (the correlation coefficient is trivial, at -0.025), suggesting that these two dimensions of teaching quality are essentially unrelated. That is, charismatic teachers who leave scores of majors in their wake, appear to be no better or worse at teaching the material than their less inspiring counterparts; teachers who are exceptional at conveying course material are no more likely than others at inspiring students to take more courses in the subject area.

These separate measures of teaching excellence allow us to address empirically whether those faculty who do particularly well in the classroom pay a price in terms of their scholarship. But first, a word about our scholarship measures. While measuring scholarly excellence is somewhat less contentious than is the case of evaluating teaching effectiveness, it is nonetheless fraught. In some fields, well-received books indicate success, in others it is performances, and still others it is highly-cited articles or the awarding of grants. How might one recognize stellar scholarship across chemistry and theater, engineering and music, economics and English, mathematics and anthropology?

We employ two very different scholarship measures. Fortunately, for each year since 1988, Northwestern has had a faculty committee comprised of distinguished professors from a wide range of disciplines whose task has been to review the scholarly accomplishments of the faculty over the previous academic year, and select a subset to be honored for their research excellence at an annual dinner. Reasons for being honored include recognition by the leading scholarly organizations in their fields such as being elected into the National Academy of Sciences, Engineering or Medicine, or into the American Academy of Arts and Sciences, receipt of prestigious fellowships such as those given by the MacArthur and Guggenheim Foundations, winning major research awards from top scholarly associations, and comparable achievements. Using this measure of research prominence, 57 percent of the 170 tenured faculty in our data set have been recognized at least once as an extraordinary scholar.

As an alternative measure, we followed a more traditional approach and constructed for each faculty member a within-department indicator of how influential

that person's scholarly work has been. Specifically, we compute a scholar's h-index—an indicator that simultaneously measures frequency of publications and the scholarly influence of those publications, thereby capturing aspects of both a researcher's productivity and the significance of that person's work. We then adjust this h-index so that we are comparing each scholar to his or her own colleagues at Northwestern. We carry out this within-Northwestern-department adjustment to take into account the fact that publication and citation norms vary dramatically across disciplines, and because some Northwestern departments are more eminent than others. Nonetheless, there exists very substantial within-department variation in tenured faculty h-indices.

These two measures of research quality are much more highly-correlated than are our two measures of teaching success: faculty members whose research have been recognized by the university average in the 49th percentile of tenured faculty field-adjusted h-indices, while those who have not been recognized for their research average in the 36th percentile of tenured faculty field-adjusted h-indices.^{xiv}

Results

With these two measures of teaching quality and two measures of research quality, we make four comparisons of teaching quality and research quality among the tenured Northwestern faculty in our sample. Our bottom line is that, regardless of our measure of teaching and research quality, there is no apparent relationship between teaching quality and research quality.

Figure 1: Relationship between percentile rank of instructor value added and probability of being recognized for research excellence

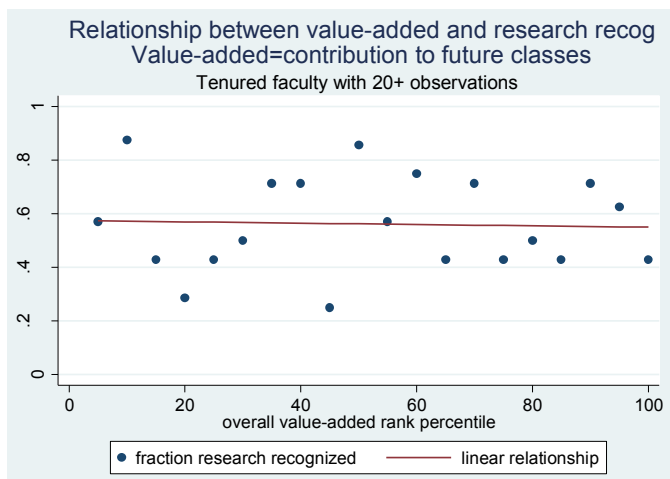
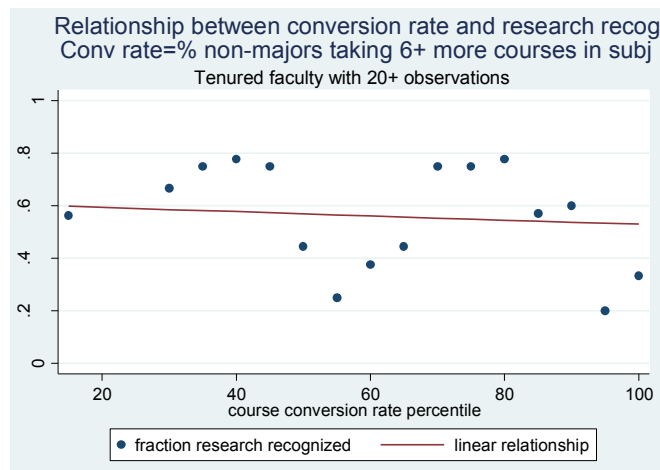


Figure 1 compares our value-added measure of teaching quality to the probability of being recognized for one's research. For ease of illustration, we group the faculty members into 20 equal-sized instructor quality bins, but we use the disaggregated data to estimate relationships. The relationship is essentially flat: with each percentile improvement in measured teaching value-added, a faculty member is 0.025 percentage points less likely to be recognized for research quality. This is a very precisely-estimated zero: the standard error of this estimate is just 0.14 percentage points. Put differently, an instructor at the 75th percentile of the instructor value-added distribution is only one percentage point less likely to be recognized for research evidence than would an instructor at the 25th percentile of the value-added distribution.

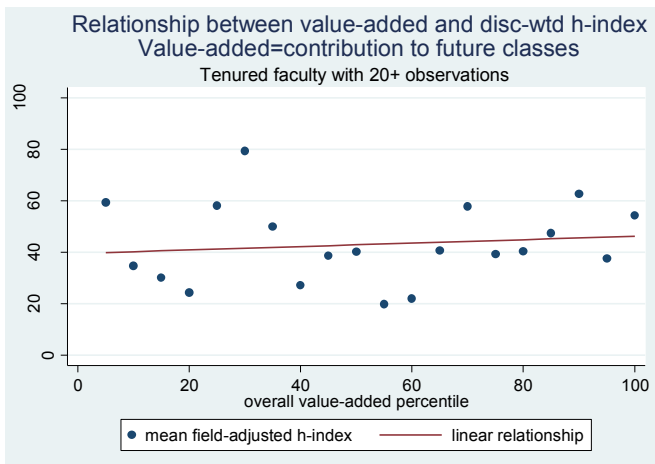
Figure 2: Relationship between percentile rank of instructor "conversion rate" and probability of being recognized for research excellence



We see the same lack of a relationship when we instead measure instructor quality using the "conversion rate"—an indicator capturing a very different aspect of instructor quality than the value-added measure does. As with the previous measure of instructor quality, we express this as a within-sample ranking. We present this relationship in Figure 2. (Note that the leftmost point on the graph is unevenly spaced because 32 percent of faculty studied convert no undecided students to majors.) We observe again that there is a precisely-estimated zero relationship between this alternative measure of instructor quality and the probability of research recognition: with each percentile increase in the instructor's conversion rate rank, a faculty member is 0.08 percentage points less likely to be recognized for research quality. The

standard error of this estimate is just 0.13 percentile points. This means that an instructor at the 75th percentile of the “conversion rate” distribution is just four percentage points less likely to be recognized for research evidence than would an instructor at the 25th percentile of the distribution.

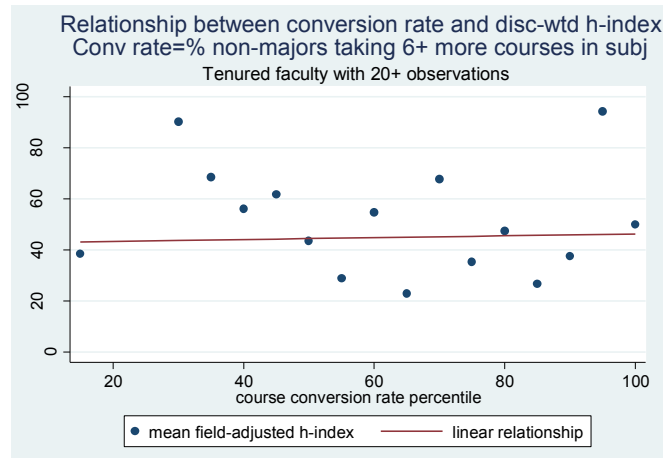
Figure 3: Relationship between percentile rank of instructor value added and percentile rank of discipline-weighted h-index



We repeat the same two analyses using the field-adjusted h-index as a measure of research excellence. Figure 3 compares our value-added measure of teaching quality to a faculty member’s percentile rank in the field-adjusted h-index. Again, the relationship is virtually flat: with each percentile point improvement in measured teaching value-added, a faculty member is 0.067 percentile points higher in the h-index ranking. The standard error of this estimate is 0.114. Therefore, the difference between the 25th and 75th percentile of the teacher quality distribution, measured in terms of value-added, is just three percentile points in the h-index distribution (and the opposite signed relationship as seen with the other measure of research quality).

Figure 4 presents the same comparison, with the “conversion rate” measure of instructor quality. With each percentage point improvement in measured teacher quality, a faculty member is 0.037 percentile points higher in the h-index ranking (standard error of 0.108), implying a difference in the h-index distribution of only two percentile points between the 25th and 75th percentile teachers.

Figure 4: Relationship between percentile rank of instructor “conversion rate” and percentile rank of discipline-weighted h-index



In sum, regardless of our measure of effective teaching or exemplary scholarship, we find that top teachers are no more or less likely to be especially productive scholars than their less accomplished teaching peers. This is encouraging for those who fear that great teachers specialize in pedagogy at the expense of research. On the other hand, it is disappointing to observe that weak undergraduate teachers do not make up for their limitations in the classroom with disproportionate research excellence.

Discussion, implications, and limitations

What does this analysis imply in terms of the growing trend of having introductory undergraduate courses being taught by non-tenure line faculty as opposed to “superstar” researchers? Our findings suggest that superb teaching does not come at the cost of diminished scholarship. Are great teachers poor scholars? Not according to our measures of teaching and research prominence. It appears that, at least in the scope of teaching by tenure line Northwestern faculty, the factors that drive teaching excellence and those that determine research excellence appear unrelated.

These findings have implications both for university administrators as well as for policymakers. Some individuals in these groups prioritize research excellence over teaching quality, while others prioritize

teaching excellence over research quality. Our analysis implies that policymakers worried about whether research efforts will come at the expense of teaching, or vice versa, should have their fears at least partially allayed. But what if state legislators take seriously our finding that while top teachers don't sacrifice research output, it is also the case that top researchers don't teach exceptionally well? Why have those high-priced scholars in the undergraduate classroom in the first place? Surely it would be more cost efficient to replace them either with untenured, lower paid, professors or faculty not on the tenure-line in the first place. That, of course, is what has been happening throughout American higher education for the past several decades. We would caution here that illustrious research faculty provide a draw for students and faculty alike. Even if their teaching isn't exceptional, their presence often is. Having them teaching freshmen sends a signal to the community that the school takes undergraduate education seriously—it isn't just research and the production of Ph.D. students that matters.

What about the recent move at the University of California towards effectively tenuring some of their full-time teaching faculty? This analysis suggests that if one of the motivations for moving undergraduate teaching from faculty with responsibility for both teaching and research to faculty whose sole responsibility is teaching is to protect the time of the former group for scholarship, this assumption needs to be questioned. Moreover, our previous paper shows that the gap in teaching performance between tenure-line and contingent faculty depends entirely on differential teaching at the low end of the value-added distribution—there are very few teaching faculty demonstrating very low levels of teaching excellence as opposed to the tenure-line faculty, where the bottom fifth or so display extremely weak teaching.^{xv} Presumably, weak contingent faculty are not renewed. While we certainly see the strong benefit of offering job security for teaching-track faculty (and recognize that higher levels of job protections likely attracts more excellent teachers to the university), giving them de facto tenure would eliminate this important lever for department chairs, deans and provosts.

Of course, this analysis has its limitations. For one, it's possible that we find no relationship between teaching success and research quality because we can measure neither perfectly, or because there's not enough variation in one measure or another to detect

results. However, the fact that we find very “precise zeros”—that is, we don't find statistically significant relationships even though we have the statistical power in our data to detect even very modest relationships—implies that neither measurement error nor a lack of sufficient variation are what's driving our inability to detect a relationship between teaching and research quality. It's also possible that our results are driven by “star” researchers teaching fewer classes than other tenured faculty, but the within-department variation in teaching loads among tenured faculty at Northwestern don't seem large enough to be responsible for our findings.

A bigger shortcoming of our analysis has to do with external validity: Northwestern occupies a rarefied place in the distribution of American universities, and it may be the case that Northwestern faculty members are particularly adept at balancing both their research and teaching expectations. We look only at tenured faculty members—so they've already experienced a very stringent screen regarding research quality, and perhaps many extraordinary teachers who were relatively weak scholars didn't make it into our analysis.^{xvi} And don't forget that research universities—and liberal arts colleges with significant research expectations for their faculty—are a modest part of all of U.S. higher education. Most professors teach heavy loads with little or no research expectations. For them, the question of whether their undergraduate teaching adversely affects their research productivity is moot. Moreover, research universities have teaching expectations that are much lower than at other schools. Some faculty at regional or local universities or colleges will undoubtedly argue that if they were teaching 3 or 4 courses per year as opposed to 8 or more, they too would have the time to be recognized scholars.

But still, research matters at places that take it seriously. The reason why most of the top-rated universities in the world are located in the United States is not what goes on in its classrooms; it is the research power of its faculties. This is reflected as well by the recent finding by Courant and Turner that faculty salaries at research universities are determined primarily by research performance and the reputation that comes with it.^{xvii} We hope that faculty and administrators at other institutions will be inspired by this study to carry out similar exercises to see whether their tenured faculty teaching entering undergraduates pay a price in terms of research productivity.

Read a college guidebook or go on a college tour. Over and over you see pictures of and hear stories about superstar research faculty teaching freshmen. Pulitzer Prize winners, Nobel Laureates, National Academy Members—all in the undergraduate classroom. Whether that properly represents reality is one question; what we address here is whether it should represent reality. Without finding a tradeoff between

great teaching and great research, we believe this to be an advantageous allocation of faculty talent.

The authors did not receive financial support from any firm or person for this article or from any firm or person with a financial or political interest in this article. They are currently not an officer, director, or board member of any organization with an interest in this article.

ⁱ Even private universities are publically supported. Their endowments grow without being subject to taxes on capital gains, and their donors receive tax deductions for their gifts.

ⁱⁱ See, for example, Mohammad Qamar uz Zaman, “[Review of the Academic Evidence on the Relationship Between Teaching and Research in Higher Education](#),” United Kingdom Department for Education and Skills, Research Report RR506, and John Hattie and Herbert Marsh, “[One Journey to Unravel the Relationship Between Research and Teaching](#),” proceedings from “Research and Teaching: Closing the Divide? An International Colloquium,” Winchester, UK, 2004, for slightly dated literature reviews.

ⁱⁱⁱ One recent example is Hee-Je Bak and Do Han Kim “[Too Much Emphasis on Research? An Empirical Examination of the Relationship Between Research and Teaching in Multitasking Environments](#),” *Research in Higher Education*, December 2015.

^{iv} See [Qamar uz Zaman](#) for a summary.

^v For examples, see Meredith Adams and Paul Umbach, “[Nonresponse and Online Student Evaluations of Teaching: Understanding the Influence of Salience, Fatigue, and Academic Environments](#),” *Research in Higher Education*, August 2012; Lillian MacNell, Adam Driscoll, and Andrea Hunt, “[What’s in a Name: Exposing Gender Bias in Student Ratings of Teaching](#),” *Innovative Higher Education*, August 2015; and Pieter Spooren, Bert Brockx, and Dimitri Mortelmans, “[On the Validity of Student Evaluation of Teaching: The State of the Art](#),” *Review of Educational Research*, October 2013.

^{vi} [Qamar uz Zaman](#).

^{vii} David N. Figlio, Morton O. Schapiro, and Kevin B. Soter, “[Are Tenure Track Professors Better Teachers?](#)” *Review of Economics and Statistics*, October 2015.

^{viii} Our findings are insensitive to our decisions

regarding cutoffs of observations needed for inclusion in the analysis. We have looked at inclusion cutoffs as low as 5 students or as high as 50 students.

^{ix} [Figlio, Schapiro, and Soter](#).

^x Eric Bettinger and Bridget Terry Long, “[Does Cheaper Mean Better? The Impact of Using Adjunct Instructors on Student Outcomes](#),” *Review of Economics and Statistics*, August 2010.

^{xi} [Figlio, Schapiro, and Soter](#).

^{xii} Scott Carrell and James West, “[Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors](#),” *Journal of Political Economy*, 2010.

^{xiii} Importantly, we find the same basic relationships for both introductory classes and upper-division/intermediate classes, though the latter relationships are estimated less precisely due to smaller sample sizes.

^{xiv} The typical tenured faculty member teaching first-year students ranks a bit below the average h-index for all tenured faculty (averaging at the 44th percentile). This is consistent with the recent finding by Courant and Turner (“[Faculty Deployment in Research Universities](#),” NBER working paper 23025, January 2017) that faculty members with higher research output teach fewer undergraduate students and undergraduate courses in research universities.

^{xv} [Figlio, Schapiro, and Soter](#).

^{xvi} But this doesn’t seem to be the case. We evaluated the data we used in [Figlio, Schapiro, and Soter](#), and found that advanced assistant professors who left Northwestern (either because they were formally denied tenure or for other reasons) were no more successful in the classroom than were advanced assistant professors who would soon receive tenure. Therefore, we suspect that this bias at least is relatively small.

^{xvii} [Courant and Turner](#).