Journal of Economic Perspectives—Volume 35, Number 1—Winter 2021—Pages 143–162

Staffing the Higher Education Classroom

David Figlio and Morton Schapiro

There are roughly 4,000 colleges and universities in the United States, including 1,600 publics, 1,700 private non-profits, and 700 private for-profits. Together, they enroll around 17 million undergraduates and another three million graduate and professional school students. Teaching is a primary mission at all of these institutions; for the vast majority of these schools, it is their singular mission. Therefore, a centrally important decision that colleges and universities face is making informed choices about how to staff their classrooms.

This decision has a number of important dimensions. Investing in a tenure-track faculty member could mean a commitment that spans several decades, and in the United States—since the 1994 change in federal law that abolished mandatory retirement at a certain age—it is a commitment with an uncertain end date. A lack of ability to forecast future demand for certain subject areas may encourage institutions to hire more contingent faculty and fewer tenure-line faculty, a decision that might affect instructional quality. Institutions also differ substantially in the degree to which the demographic characteristics of their faculty match those of their student bodies, and this might also affect students' outcomes. Moreover, while we know that teaching quality is massively heterogenous across the categories of potential instructors, we also know that it is extremely challenging to

■ *David Figlio is Orrington Lunt Professor and Dean of the School of Education and Social Policy and Morton Schapiro is President and Professor of Economics, both at Northwestern University, Evanston, Illinois. Their email addresses are figlio@northwestern.edu and nu-president@northwestern.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.35.1.143.

measure teaching effectiveness. As we discuss below, student evaluations of classroom teaching are subject to considerable bias and room for manipulation.

One objective of this paper is to explore different ways to evaluate what students experience in the undergraduate classroom. A central question is whether there is a more objective way to evaluate a teacher than through the ubiquitous course evaluation survey. We develop an alternative approach that seeks to measure what students actually learn. Once we do that, we address a secondary question: whether especially charismatic teachers—defined as those who are particularly adept at inspiring students to major in their disciplines—are better (or worse) at having their students learn than their less captivating peers.

The relatively small number of institutions that pride themselves not just on their teaching output but also on the scholarly contributions of their faculty face a multi-tasking problem of the first order. Too much emphasis on research may leave their students at all levels—undergraduate, graduate, and professional—to fend for themselves. While there appears to be important peer effects among students attending selective institutions so that a good deal of learning occurs outside the classroom, one would hope that transformative learning is also taking place within the classroom. On the other hand, too much focus on teaching might mean that the self-described "research" university is that in name only.[1]

While we don't pretend to offer a magic formula for determining the ideal distribution of resources and focus between teaching and research, we present some clues about the relationship between the two. One question we explore, at least tentatively, is whether the most accomplished teachers sacrifice in terms of their scholarship, and, alternatively, whether relatively poor teachers make up for it in research excellence. Throughout this paper, we review the literature and then offer empirical results based on students and faculty at our own university, Northwestern. Whether the findings are generalizable to other schools is an open question that we hope others will be encouraged to explore.

[1] In addition to teaching and research, many prominent universities are also in the entertainment business. Their plays and concerts might occasionally attract attention beyond their local communities, but if they participate in "big time" sports, national and even global attention are virtually guaranteed. For example, when Northwestern played Ohio State for the Big Ten conference football championship in December 2018, it was viewed in 8.7 million households. While it is commonplace for major public research universities to participate at the highest level of athletics, it is unusual for their private counterparts to do the same. Of the 27 private universities that are members of the prestigious Association of American Universities, only five are among the 65 members of the athletic "power conferences": Stanford and USC in the Pac 12, Vanderbilt in the SEC, Duke in the ACC, and Northwestern in the Big Ten. Most of the rest are either in the Ivy League or compete in the NCAA Division III (which prohibits athletic scholarships).

Measuring Teaching Effectiveness

While it is commonplace to have students fill out evaluation forms at the conclusion of the semester or quarter, a large literature shows that the results appear to be biased by gender, race, and nationality. White American men are often given higher ratings than others and, without objective measures of student learning, it is impossible to evaluate whether those ratings are actually "earned." Mengel, Sauermann, and Zoelitz (2019) study a context at Maastricht University in which students are randomly assigned to male or female section instructors, and show that students systematically rate female instructors worse—a pattern driven by evaluation of male students. Despite these biases, there may be some value in student evaluations: Hoffmann and Oreopoulos (2009a), for example, provide evidence using administrative data from a large Canadian university that higher student perception of professor quality is associated not just with a reduction in course-dropping, but also in increased performance in that class. But this signal may be a noisy one: Deslauriers et al. (2019) examine Harvard undergraduates enrolled in large introductory physics classes and find that "attempts to evaluate instruction based on students' perceptions of learning" could be very misleading, misstating the actual learning (as determined by multiple choice tests) that takes place. They warn that "a superstar lecturer could create such a positive feeling of learning that students would choose those lectures over active learning" despite the fact that, according to their analysis, students in "active" classrooms actually learned more.[2]

Concerns about bias have led the American Sociological Association (2019) to caution against over-reliance on student evaluations of teaching, pointing out that "a growing body of evidence suggests that their use in personnel decisions is problematic" given that they "are weakly related to other measures of teaching effectiveness and student learning" and that they "have been found to be biased against women and people of color." The ASA suggests that "student feedback should not be used alone as a measure of teaching quality. If it is used in faculty evaluation processes, it should be considered as part of a holistic assessment of teaching effectiveness." Seventeen other scholarly associations, including the American Anthropological Association, the American Historical Association, and the American Political Science Association, have endorsed the ASA report (for discussion in the education press, see Flaherty 2019b; Supiano 2019). For all of these reasons, it's clear that student evaluations are far from excellent summative assessments of instructor effectiveness.

An alternative method to evaluate teaching is to examine whether faculty induce student interest in a subject, as measured by the likelihood that students take additional courses in that subject (Bettinger and Long 2010; Hoffmann

[2]Some other studies of student evaluations of instructional quality include Bavishi, Madera, and Hebl (2010), Beleche, Fairris, and Marks (2012), Boring (2017), Boring, Ottoboni, and Stark (2016), Braga, Paccagnella, and Pellizzari (2014), Rivera and Tilcsik (2019), Stark and Freishtat (2014), and Stroebe (2016). Flaherty (2016) provides a useful summary of some of the literature.

and Oreopoulos 2009a). While it is certainly reasonable to assume that teachers who inspire students to continue studying their discipline are especially good at conveying knowledge, that need not be the case. Charisma might not translate into teaching effectiveness. An instructor who increases the likelihood that students take additional classes in a subject may or may not produce "deep learning," in the words of Carrell and West (2010). One could imagine that charismatic instructors could even inhibit deep learning if they induce students into thinking that their *enjoyment* of a subject implies *mastery* of that subject.

Can we actually measure learning? One way to get a handle on this question, at least in part, would be to examine (with appropriate statistical controls) how students perform in subsequent classes. If faculty members are especially effective at teaching, say, introductory microeconomics, we would expect their students to perform unexpectedly well in intermediate micro and beyond. Perhaps those faculty might not be particularly good at generating majors, and they might not have the best teaching evaluations, but being a "popular" teacher and being a "successful" one are not necessarily the same thing. Carrell and West (2010) pioneered a measure of professor quality based on academic performance in subsequent courses within the same subject, a method continued by Figlio, Schapiro, and Soter (2015) and Ran and Xu (2019).

We are curious about whether a relationship exists between charismatic or inspirational teaching and the "deep learning" reflected by value added to subsequent classes. As an exploratory first step, in this paper we make use of 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. Specifically, 170 tenured faculty members taught at least 20 first-quarter students across the eight annual cohorts. (We focus on tenured faculty because we also want to compare teaching effectiveness measures with research quality measures.) While Northwestern is an admittedly unrepresentative institution, our analysis should at least offer some clues regarding where and how to look at other colleges and universities.

We construct two measures of teaching quality. First, we look at the ability of professors to convert students into majors. A talented chemistry teacher in an introductory course may lead a declared chemistry major to keep that major, or 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 attract students to the major presumably reflects one dimension of teaching excellence. Second, we consider a "deep learning" measure described in considerable detail in Figlio, Schapiro, and Soter (2015). We look first at the likelihood of taking additional courses in a subject area, and then measure the deviation in the grade received by a student in follow-up courses in that subject. A successful undergraduate teacher in, say, introductory psychology, not only induces students to take additional psychology courses but leads those students to do unexpectedly well in those additional classes (based on what we would have predicted given their standardized test scores, other grades, grading standards in that field, and so on). In our 2015 paper, we lay out the

statistical techniques employed in controlling for course and student impacts other than those linked directly to the teaching effectiveness of the original professor.[3]

In essence, therefore, our two measures of teaching quality reflect, in the first case, inspiration, as indicated by the ability to convert students to a subject that they had not previously planned on studying in-depth and in the second case, value added (or "deep learning"), that is transferrable to subsequent classes in the subject.

There is considerable range in these variables: The 25th percentile tenured professor ranked according to the "inspiration" measure converted no students to the major in our data, while the 75th percentile tenured professor converted 25 percent of students to the major, and the 90th percentile tenured professor converted 46 percent of students to the major. When ranked according to "deep learning," the difference between the 90th and 10th percentile tenured professor is over one-tenth of a letter grade, and the difference between the 75th and 25th percentile tenured professor is over one-twentieth of a letter grade.

Are More Charismatic Teachers Also Better Teachers?

How are these two measures of teaching success related? Are charismatic professors who attract students to the major also unusually proficient at imparting knowledge that improves performance in future courses? The study by Deslauriers et al. (2019) of Harvard undergraduates in a physics class suggests they might actually be worse. This concern was expressed in an account of their work that appeared in the education press, with the provocative title: "Study: How Smooth-talking Professors Can Lull Students Into Thinking They've Learned More Than They Have" (Flaherty 2019a).

Based on our analysis of the 170 tenured Northwestern faculty, we find that those who are most successful in inspiring students to become majors in their subject are not any more distinguished in facilitating "deep learning" than their less charismatic counterparts: 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, teachers who leave scores of majors in their wake appear to be no better or worse at teaching the material needed for future courses than their less inspiring counterparts; teachers who are exceptional at conveying course material are no more likely than others to

[3]While this measure of teaching excellence goes far beyond course evaluations, it is of course 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 dataset centers on success in teaching courses open to first-quarter freshmen, most of which are introductory courses. Perhaps some faculty are poor teachers for an intro course aimed at first-year undergraduates but do a better job teaching senior seminars for undergraduates or graduate courses. That said, this measure at least provides a (selected) analog to the "value added" measurements of K–12 instruction.

inspire students to take more courses in the subject area. We would love to see if this result would be replicated at other institutions.

Of course, attracting majors may be a reward in itself, at least at the departmental level, should more majors translate into additional faculty slots and more funding at the departmental level. At the institutional level, however, choice of major is basically a zero-sum game, as adding one major typically means losing a different one (unless a student decides to double or triple major). There may also be a zero-sum aspect of our measure of "deep learning" as well: If a faculty member inspires extra studying in one subject, does that reduce student attention to other subjects? Of course, the reverse might be true as well, if a particularly effective instructor helps foster deeper learning in other subjects. Unpacking this question could be a fruitful avenue for future research.

Is There a Tradeoff between Research and Teaching Excellence?

Our two measures of teaching excellence described above 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 measures of scholarship.

While measuring scholarly excellence is somewhat less contentious than evaluating teaching effectiveness, it is nonetheless fraught. In some fields, well-received books indicate success, in others it is in-person 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?

As described in Figlio and Schapiro (2017), we employ two very different scholarship measures. For each year since 1988, Northwestern has had a faculty committee comprised of distinguished professors from a wide range of disciplines who review the scholarly accomplishments of the faculty in the previous academic year, and select a subset of faculty 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. The committee does also make certain to identify appropriate awards at all career stages, so more junior scholars winning a "rising star" award from a top professional society or a National Science Foundation CAREER grant would also make the cut for inclusion. Using this measure, 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

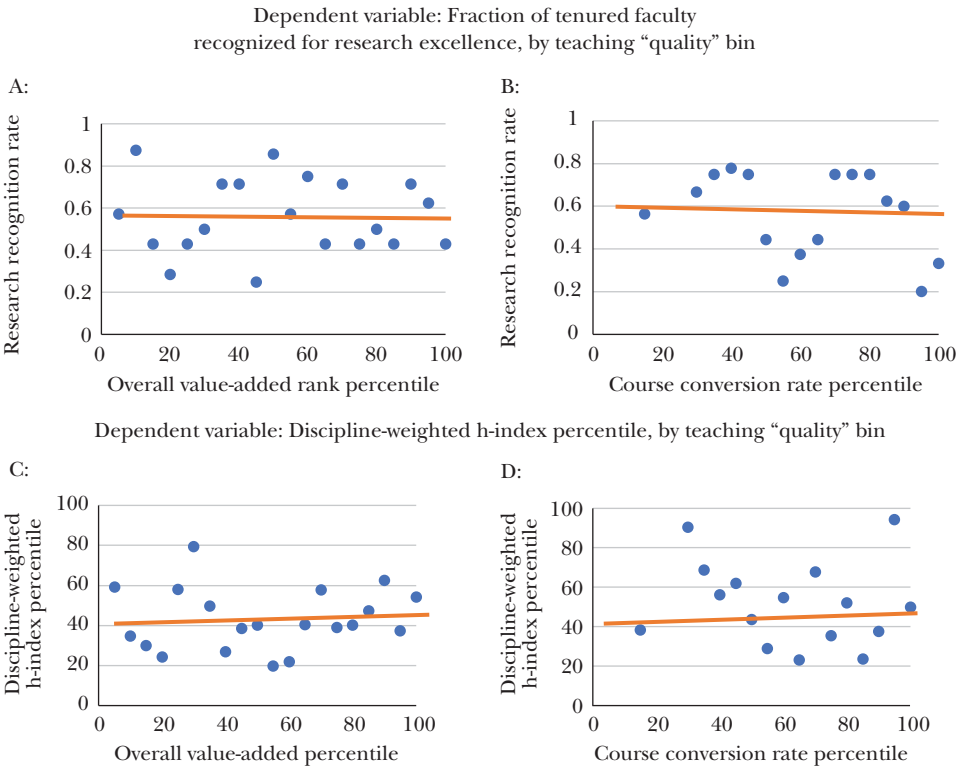
h-index—an indicator that simultaneously measures frequency of publications and the scholarly influence of those publications, thereby capturing aspects of both a researcher's productivity and the significance of that person's work. The h-index is defined as the number of articles or papers with at least *h* citations: For example, a professor with 20 papers that had been cited at least 20 times would have an h-index of 20. We adjust this h-index so that we are comparing scholars only to colleagues within their own department at Northwestern; specifically, we standardize h-indices within each academic department of 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—among tenured Northwestern faculty, the typical 75th percentile department-adjusted h-index is 1.5 standard deviations higher than the typical 25th percentile department-adjusted h-index, and the typical 90th percentile department-adjusted h-index is 2.4 standard deviations higher than the typical 25th percentile department-adjusted h-index—as one would expect given that this measure of scholarly work has some bias toward older and more established faculty members. (The first measure may also be modestly biased toward more established faculty members but not nearly as much as the h-index is.)

These two measures of research quality are much more highly correlated than our two measures of teaching success: Those Northwestern faculty members teaching introductory first-year courses whose research has been recognized by the university average in the 49th percentile of tenured faculty field-adjusted h-indices, while those tenured introductory course teachers who have not been recognized for their research average in the 36th percentile of tenured faculty field-adjusted h-indices. The fact both groups of faculty are at or below the university median suggests that introductory classes are disproportionately taught by faculty with somewhat less research success, at least by this measure. This last point, however, could be due in part to the previously mentioned bias of h-indices toward more senior faculty members, given that very senior faculty members are somewhat less likely to teach introductory courses.

With these two measures of teaching quality and two measures of research quality, we make four comparisons in Figure 1 of teaching quality and research quality among the *tenured* Northwestern faculty in our sample. Our bottom line is, regardless of our measure of teaching and research quality, there is no apparent relationship between teaching quality and research quality.

In Figure 1, panel A compares our value-added measure of teaching quality for future courses 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

Figure 1
Relationship between Measures of Teaching Excellence and Measures of Research Excellence



Source: This figure was previously published in Figlio and Schapiro (2017).
Note: Data include tenured faculty at Northwestern with 20 or more student observations. Instructor "value-added" is measured as the contribution to future course letter grades in the subject, calculated using the method developed by Figlio, Schapiro, and Soter (2015). The "course conversion rate" is the percentage of non-majors who ultimately take six or more subsequent courses in the subject.

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.

In Panel B, we see the same lack of a relationship when we instead measure instructor quality using a "conversion rate"—the fraction of initially non-majors who ultimately take six or more courses in the subject. As with the previous measure of instructor quality, we express this as a within-sample ranking. (The leftmost point on the graph is unevenly spaced because 32 percent of faculty studied convert zero undecided students to majors.) Again, we observe 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.

We repeat the same two analyses using the field-adjusted h-index as a measure of research excellence. Panel C 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).

Finally, Panel D 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.

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.

So what does this analysis imply in terms of staffing the undergraduate classroom? 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. Of course, it's possible that this result arises because teaching or scholarship are imperfectly measured, but we are finding very "precise zeros" between variables with a reasonably high level of variation—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. At least in the scope of teaching by tenure-line Northwestern faculty, the factors that drive teaching excellence and those that determine research excellence are unrelated.

These findings have implications for university administrators and 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 in the classroom either with untenured, lower-paid professors, or with 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, as we discuss in detail in the section that follows. And, of course, there's the other potentially uncomfortable question that our analysis implies: Should we be concerned about the possibility that the weakest scholars amongst the tenured faculty are no more distinguished in the classroom than are the strongest scholars? Should expectations for teaching excellence be higher for faculty members who are on the margin of tenurability on the basis of their research excellence?

Should the Move toward Contingent Faculty in US Universities Concern Us?

The role of tenure in American higher education has declined dramatically in recent decades. In 1975, 57 percent of all faculty (full-time and part-time, excluding graduate students) were in the tenure system; by 2011, the proportion was about half that high at 29 percent, and it has continued to fall since. The American Association of University Professors website presents its Contingent Faculty Index summarizing data from the IPEDS Fall Staff Survey, and *The Chronicle of Higher Education* regularly summarizes the latest numbers. For example, Simonton (2019) reports that non-tenure-track appointments grew from 10 percent of all full-time faculty positions in 2008–09 to 27 percent in 2018–19 at the 870 institutions that participated in the American Association of University Professors' Faculty Compensation Survey in both years, and that 56 percent of full-time and part-time faculty members at four-year public institutions and 66 percent of those at four-year private nonprofit institutions were not on the tenure track in 2017. At two-year public colleges, fewer than 20 percent of faculty are in the tenure system.

There has been ongoing speculation in the education press about the level at which this decline in tenure might bottom out. A common prediction is that the share of tenure track/tenured faculty will eventually stabilize at between 15 to 20 percent, with tenure being largely limited to the flagship public and private research universities and the wealthiest of the liberal arts colleges (Wilson 2010). Morson and Schapiro (2015) predict that by 2040, only around 10 percent of faculty positions will be held by tenure-track/tenured professors.

This reduction in tenure trend seems to have accelerated since January 1, 1994, when the mandatory retirement age for college and university faculty was abolished by federal law. As a result, colleges that wanted an older faculty member to retire

have often sought to negotiate multi-year plans with incentives for a "phased retirement"—but a faculty member who does not wish to sign an agreement for such a plan cannot be required to do so. As a result, granting tenure involves additional end-of-career uncertainties, and shifting more instruction to contingent faculty mitigates those risks. Especially notable is the rise of the full-time, contingent faculty members at PhD-granting universities. Their representation within the entire group of full-time faculty went from 24 percent to 35 percent at public doctoral institutions from 1995 to 2011 and from 18 percent to 46 percent at private nonprofit doctoral institutions (Ehrenberg 2012).

The erosion of tenure has raised a number of concerns. Some observers have lamented the potential blow to academic freedom dealt by the decline of tenure, while others have focused on the often-challenging employment conditions under which many contingent faculty work (for example, June 2012; Wilson 2010). Further, McPherson and Schapiro (1999) point to efficiency gains from tenure; they outline its positive role in influencing the distribution of authority within colleges and universities. Here, we focus on educational outcomes: Do undergraduates taught by contingent faculty members learn as much as those taught by faculty who are tenured or on a tenure-track appointment?

On a national level, Ehrenberg and Zhang (2005) present evidence that hiring more part-time and contingent faculty lowers institutional graduation rates. This result is bolstered by Bettinger and Long (2010), who find a similarly negative effect on aggregate levels of persistence when they focus specifically on part-time adjuncts. These types of results indicate that even if contingent faculty are more popular with students—perhaps because of classroom behaviors that maximize student evaluations but not student learning—they nonetheless might not be successful in improving students' longer-term prospects.[4]

Several recent papers have made substantial steps toward understanding whether tenure-line faculty members outperform (or perform worse than) contingent faculty in the classroom. Bettinger and Long (2010) study this question using data largely centered on data from twelve public four-year colleges in the state of Ohio whose principal purpose is teaching. Their creative identification strategy involves treating short-term vacancies in departments as random events, and thus essentially analyzing the effects on learning from transitory adjunct faculty. While their analysis therefore may not speak to the effects of part-time faculty with longer-term contracts, and certainly does not address the effects of full-time contingent faculty, it does reflect the types of "one-off" contingent faculty hired with some regularity at many institutions. Bettinger and Long find some evidence that contingent faculty induce student interest in a subject, as measured by the likelihood that students take additional courses in that subject, though their data do not provide the opportunity to study how students perform in subsequent classes. Ran and Xu (2019), studying both two-year and four-year

[4] Carrell and West (2010) show that instructors who have better student evaluations tend to produce lower levels of "deep learning."

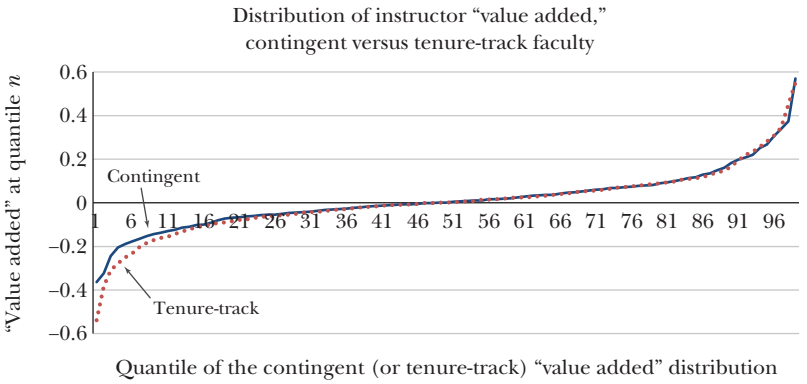
teaching-oriented colleges in an anonymous state university system, measure both "deep learning" in the sense of performing better in future courses à la Carrell and West (2010) and the likelihood of subsequent course-taking à la Bettinger and Long (2010). They find that contingent faculty perform worse on both measures than do tenure-line faculty. On the other hand, Feld, Salamanca, and Zoelitz (2020) study students randomly assigned to tutorials in a Dutch business school and find negligible differences across instructor type in effects on students' current and future academic outcomes, job satisfaction, and earnings. In sum, while the evidence varies from setting to setting, it is certainly not obvious that tenure-line faculty members are more effective instructors in teaching-intensive settings.

What about results specific to research universities, where the aforementioned multi-tasking problem of maximizing an objective function that includes both the production of cutting-edge research and the provision of outstanding undergraduate teaching may be more pronounced? Hoffmann and Oreopoulos (2009a) were the first to evaluate the instructional quality of tenure-line versus contingent faculty in a research university setting, though the Canadian research university context they study only permits analyzing the likelihood that students take additional classes in the same subject rather than observe their academic performance in future classes. They find no evidence that contingent faculty are either better or worse at inspiring students to take more classes in their subjects.

In Figlio, Schapiro, and Soter (2015), we carry out an analysis at Northwestern, where we also can measure both performance in future related courses and the likelihood of subsequent course-taking. While Northwestern contingent faculty do not have the employment protections that many at places like the University of California (which has a status of "Lecturer with Security of Employment") have, Northwestern is still unusual relative to most institutions in that contingent faculty members tend to have stable, long-term relationships with the university and a substantial majority are full time. (Generally, these are renewable three-year contracts, and most of the contingent faculty in our study are in their second or later contract.) Thus, our results should be viewed in the context of where non-tenure faculty at a major research university function as designated teachers (both full-time and part-time) with long-term relationships to the university. We find that, on average, tenure-line faculty members do not teach introductory undergraduate courses as well as do their (largely full-time, long-term) contingent faculty counterparts. In other words, our results suggest that on average, first-term freshmen learn more from contingent faculty members than they do from tenure track/tenured faculty.

Are our results driven by a handful of outliers or a by larger swath of the distribution? That is, do these differences arise because most tenure track/tenured faculty members perform similarly to most contingent faculty members? Or are the differences due to events at the tails of the distribution, with either the best contingent faculty teachers substantially outperforming the best tenure track/tenured teachers or the worst tenure track/tenured teachers performing considerably worse than the worst contingent faculty teachers?"

Figure 2
Distribution of Instructor "Value Added," Contingent versus Tenure-Line Faculty



Source: This figure was first published in Figlio, Schapiro, and Soter (2015).

To explore this question, Figure 2 compares the distributions of value-added in future related courses of individual contingent faculty teachers and tenure track/tenured teachers. In the figure, we plot a variant of the cumulative density function where the percentile in each distribution is on the horizontal axis and the corresponding value-added measure is on the vertical axis, which makes clear what part of the distribution is generating the results. An individual instructor's value-added is an instructor-specific fixed effect retained from our preferred specification (in which we estimate instructors' effect on grade points earned in the next course, controlling for both student fixed effects and next-course fixed effects). As can be seen in the figure, the top three-quarters of the contingent faculty and tenure track/tenured faculty distributions are virtually perfectly overlapping, so the most outstanding contingent faculty members and most outstanding tenure track/tenured faculty members perform essentially identically, and the same is true at other points in the distribution such as the median. But the bottom quarter of the tenure track/tenured faculty have lower value-added than the bottom quarter of the contingent faculty, and this difference is substantial for the bottom 13 percent of the distribution (around the weakest 150 instructors, by our definition, amongst those who taught introductory courses over the decade of time covered in our study). It is clear that our results are not being driven by a handful of outliers, but it is also clear that the difference in average outcomes is due to the differences at the bottom of the value-added distribution.

In some ways, this pattern is exactly what we might have expected: Contingent faculty members who are hired to teach and who perform relatively poorly are less likely to be renewed than are those who perform well, while tenure-track faculty who are relatively poor teachers may be promoted and retained for reasons other than their teaching ability. But are there specific differences between contingent faculty members and tenure track/tenured faculty members that can explain our findings? To shed some light on this issue,

we collected curriculum vitae available through extensive web-searching. Two directly observable variables are years of experience (calculated based on time since PhD and employment history) and native language (calculated based on the country in which a faculty member earned his or her bachelor's degree or its equivalent). Tenure track/tenured faculty are modestly more likely to have attended undergraduate institutions in English-speaking countries (86.3 percent versus 79.2 percent for contingent faculty) and average dramatically more experience (21.9 years versus 11.6 years for contingent faculty). But these variables explain no more than a modest fraction of our results.

Rather, it is apparently the case that the bottom tail of the contingent faculty distribution are considerably more effective classroom instructors than are those at the bottom tail of the tenure-line faculty distribution, at least in the first-term freshman classroom. There are extraordinarily good reasons for top research universities facing multi-tasking problems to recruit, retain, and reward faculty members on the basis of research, but apparently this comes at the cost of having a fraction of the distribution of tenure-line faculty be disproportionately poor performers in the classroom—both in terms of inspiration and in terms of preparation for future courses.

It is, of course, also noteworthy how closely the top three-quarters of the tenure-line and contingent faculty distributions track one another in Figure 2. Most tenure-line faculty teaching freshmen at Northwestern perform just as well, according to our measures of teaching effectiveness, as do those who are recruited, retained, and rewarded on the basis of their teaching. This must be reassuring to those who worry that there's a stark tradeoff between teaching and research quality. (The results we present above, in which we look just at tenured Northwestern faculty, should also provide reassurance on this front.) Rather, the instructional quality gap between tenure-line and contingent faculty members is concentrated in a relatively small fraction of the teaching force, at least at Northwestern.

What about our other measure of teaching quality—the ability of charismatic faculty to inspire students to major in their subject? Contingent faculty members are modestly more likely to convert students into majors; the typical contingent faculty member converts 1.6 percent more students to majors than does the typical tenure-line faculty member.

An obvious question is the degree to which these findings can be generalized. Because a key part of our identification strategy is to limit our analysis to first-term freshman undergraduates, the evidence that contingent faculty produce better outcomes may not apply to more advanced courses. Further, Northwestern is among the most selective and highly ranked research universities in the world, and its ability to attract first-class contingent faculty may be different from that of most other institutions.

It is also important to note that a substantial majority of contingent faculty at Northwestern are full-time faculty members with long-term contracts and benefits, and therefore may have a stronger commitment to the institution than some of their contingent counterparts at other institutions. Northwestern's tenure track/tenured faculty members may also have different classroom skills from those at other schools.

Finally, Northwestern students come from a rarefied portion of the preparation distribution and are far from reflective of the general student population in the United States. That said, our results are strongest for the students and subjects that are most likely to generalize to a considerably wider range of institutions: The benefits of taking courses with contingent faculty appear to be stronger for the relatively marginal students at Northwestern (although they are still very well-prepared students), and our results are similar for top-ranked departments as for lower-ranked departments.

There are many aspects relating to changes in the tenure status of faculty, from the impact on research productivity to the protection of academic freedom. But certainly, learning outcomes are an important consideration in evaluating whether the observed trend away from tenure track/tenured toward contingent faculty is good or bad. Our results, coupled with that of others, provide evidence that the rise of full-time designated teachers at US colleges and universities may be less of a cause for alarm for the quality of teaching than many assume. Of course, it is important to note that our analysis is necessarily a partial-equilibrium one: While our analysis tells us on the margin how faculty types compare, it does not tell us whether the current allocation of faculty slots to contingent versus tenure-track faculty is efficient. For one thing, because contingent faculty members teach more classes per year than tenure-track faculty members, the implication that contingent faculty manage to maintain comparable (or superior) teaching with heavier teaching loads speaks to the potential tradeoffs between staffing the undergraduate classroom with more research-intensive versus more teaching-intensive faculty members.

In What Ways Does Instructor Gender, Race, and Ethnicity Matter?

In fall 2017, 11.6 percent of US college and university faculty (and 10.8 percent of tenure-line faculty) were Black, Hispanic, Native American, or mixed-race, despite 35 percent of the student body being one of these demographic groups (US Department of Education 2018; Tables 306.10 and 315.20). Also in fall 2017, 46.3 percent of faculty (and 43.2 percent of tenure-line faculty) were women, at a time when 56.7 percent of the student body were women. Large mismatches between student and faculty demographics like this have a number of consequences and risks: a lack of diversity in substantive representation, the potential for limiting the range of what is taught and how it is taught, and how it might affect major and career choices of students due to implicit stereotypes (Reuben, Sapienza, and Zingales 2014) or lack of role models (Zafar 2013).

A large literature in K–12 education suggests that outcomes such as test scores, attendance, and suspension rates are affected by the demographic match between teachers and students (Dee 2004; Dee 2005; Egalite, Kisida, and Winters 2015; Gershenson, Holt, and Papageorge 2016; Lim and Meer 2017; Lindsay and Hart 2017). This relationship appears to have a long reach: Gershenson et al. (2018) find that if a Black male student has at least one Black teacher in the third, fourth, or

fifth grade, he is significantly less likely to drop out of high school and more likely to aspire to attend a four-year college, as proxied by taking a college entrance exam, and that these effects are particularly pronounced for economically disadvantaged Black male students. Looking at data from schools in Israel, Lavy and Sand (2018) demonstrate the long-run effects on later courses, career choice, and earnings when primary teachers show gender bias: there is reason to believe that the same is true regarding race and ethnicity.

Does demographic match matter in college as well? Some recent evidence suggests that the answer is "yes." Using data from Canadian research universities, Hoffmann and Oreopoulos (2009b) show that college students are more likely to complete a course and to perform better in that course when they have a same-gender professor. Carrell, Page, and West (2010) show that having a female professor in a course in science, technology, engineering, or mathematics will substantially increase the likelihood that females will take more courses and eventually graduate with a degree in these fields. Fairlie, Hoffmann, and Oreopoulos (2014) demonstrate, using data from a large community college in California, that racial and ethnic minority faculty members reduce the minority achievement gap in class performance and dropout rate, while Kofoed and McGovney (2019) find that same-gender and same-race role models at the US Military Academy have a strong effect on occupational choice. Mansour et al. (2020) examine students at the US Air Force Academy and find that high-ability female students who were assigned a female professor had a substantial increase in the probability of working in a science, technology, engineering, or mathematics occupation and in the probability of receiving a master's degree in those fields. It seems likely, therefore, that hiring a racially, ethnically, and gender diverse faculty will benefit students in an environment, such as the current state of affairs in US higher education, where the student body is much more diverse than is the faculty responsible for teaching them.

Women and minorities are underrepresented not just among faculty members in general, but in particular among tenure-line faculty members. While Black, Hispanic, Native American, and mixed race faculty members represented 12.7 percent of assistant professors in fall 2017 and 11.6 percent of associate professors, they are only 8.2 percent of full professors (US Department of Education 2018). The gradient is even more striking for women, who comprised 51.8 percent of assistant professors—almost, if not quite, the female proportion of the student body—but 45.4 percent of associate professors and 32.8 percent of full professors in fall 2017. Women represented 49.4 percent of contingent faculty members in fall 2017, while Black, Hispanic, Native American, and mixed race individuals comprised 11.4 percent of contingent faculty. Looking at the intersection of gender and minority groups, 7.3 percent of assistant professors are women who are Black, Hispanic, Native American, or mixed race, but only 5.8 percent of associate professors and 3.2 percent of full professors are.

The role-model issue is relevant not just for students but for junior faculty as well: While in fall 2017 there were 1.79 white male full professors for every white male assistant professor, there was just 0.39 Black female full professors for every Black female assistant professor, 0.54 Hispanic female full professors for every

Hispanic female assistant professor, 0.75 Native American female full professors for every Native American female assistant professor, and 0.36 mixed race female full professors for every mixed race female assistant professor. Lundberg and Stearns (2019) identify a number of challenges that women in economics face and point to institutional policies and promotion and tenure processes that are biased against women. The disproportionate impacts of the current COVID-19 pandemic on female academics, and especially female academics of color, seem likely to further exacerbate gender inequities in the academy, and universal remedies like blanket tenure clock extensions might have the unintended consequence of disadvantaging women in the academy.[5] Of course, the issue starts before women become faculty members; for example, there exists considerable heterogeneity in the success of female economics graduate students across PhD programs (Boustan and Langan 2019).

The causes and specific consequences of these racial and gender differences in faculty rank are beyond the scope of this paper, but they signify a structural challenge that colleges and universities face. Two recent papers published in this journal, Buckles (2019) and Bayer, Hoover, and Washington (2020), provide advice for ways to make the economics profession more inclusive of and hospitable to female scholars and scholars of color at all career stages. However, the shift away from tenure-line faculty toward contingent faculty means that there are fewer opportunities to diversify the tenure ranks, putting several institutional objectives in direct tension with one another.

Conclusion

With non-tenure line faculty providing unusually good undergraduate teaching, and within the tenured group, top scholars neither better nor worse in teaching than their colleagues, why would anybody ever allocate the best senior scholars to undergraduate classes? Clearly, it would be more efficient to replace them with less expensive non-tenure eligible teachers or with younger non-tenured tenure-line faculty. But before leaping to this conclusion, we offer a few words of caution. Illustrious research faculty provide a draw for students and faculty alike. Even if their undergraduate teaching isn't exceptional, their presence often is. Having outstanding scholars teaching first-year students sends a signal to the community that the school takes undergraduate education seriously—that it isn't just research and the production of PhD students that matters. Meanwhile, while it may be tempting to hire female and minority faculty members into teaching-track positions in order to quickly provide more students with "a professor like me," the segregation of "professors like me" into lower-prestige teaching positions may reinforce stereotype threats and have deleterious effects on minority students.

[5]As one important example that is highly salient at the present moment, Antecol, Bedard, and Stearns (2018) show that gender-neutral tenure clock-stopping policies associated with parenting reduced female economists' tenure probabilities but increased male economists' tenure probabilities.

What about the recent move at the University of California and elsewhere towards effectively tenuring some of their full-time teaching faculty? Our analysis suggests that if the motivation 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 approach needs to be questioned. We have shown that, at least at Northwestern, the gap in teaching performance between tenure-line and contingent faculty depends entirely on the weakest teachers among the tenured professors. 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 attract more excellent teachers to the university), giving de facto tenure to this group might reduce the power of this important lever for department chairs, deans, and provosts. On the other hand, if doing so permits universities to invest more in research faculty members' research time, this could help them to partially solve the university's multi-tasking problem but at the potential cost of reinforcing a two-tiered faculty system, albeit with more job protections for the less-prestigious tier of faculty.

Finally, a word about the role of research at those universities that take special pride in their scholarly output. The reason why most of the top-rated universities in the world are located in the United States is not what goes on in their classrooms; it is the research power of their faculties. Read a college guidebook or go on a college tour at a top research university. Over and over, you will see pictures of and hear stories about superstar research faculty including Pulitzer Prize winners, Nobel Laureates, National Academy members—all in the undergraduate classroom and often teaching first-year students. Whether those pictures properly represent reality is one question; what we address here is whether it should represent reality. Given that we do not find a tradeoff between great teaching and great research, we believe that having top research faculty do a share of the undergraduate and first-year teaching is an advantageous allocation of faculty talent.

The multi-tasking challenge faced by our nation's most prestigious research universities isn't easy to solve. But measuring both teaching and research outcomes seems to be a useful first step and could be helpful in cracking the nut of how to staff the higher education classroom. One key lesson from our analysis is that teaching-intensive institutions and research universities alike could be well-served to pay less attention to numerical student evaluations and instead work to develop alternative ways to measure and reward good teaching. Our analysis provides some ideas about ways institutions might try to go about doing this.

References

American Sociological Association. 2019. "Statement on Student Evaluations of Teaching." *American Sociological Association*, September. https://www.asanet.org/sites/default/files/asa_statement_on_student_evaluations_of_teaching_feb132020.pdf.

Antecol, Heather, Kelly Bedard, and Jenna Stearns. 2018. "Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies?" *American Economic Review* 108 (9): 2420–41.

Bavishi, Anish, Juan Madera, and Michelle Hebl. 2010. "The Effect of Professor Ethnicity and Gender on Student Evaluations: Judged before Met." *Journal of Diversity in Higher Education* 3 (4): 245–56.

Bayer, Amanda, Gary A. Hoover, and Ebonya Washington. 2020. "How You Can Work to Increase the Presence and Improve the Experience of Black, Latinx, and Native American People in the Economics Profession." *Journal of Economic Perspective* 34 (3): 193–219.

Beleche, Trinidad, David Fairris, and Mindy Marks. 2012. "Do Course Evaluations Truly Reflect Student Learning? Evidence from an Objectively Graded Post-Test." *Economics of Education Review* 31 (5): 709–19.

Bettinger, Eric P., and Bridget Terry Long. 2010. "Does Cheaper Mean Better? The Impact of Using Adjunct Instructors on Student Outcomes." *The Review of Economics and Statistics* 92 (3): 598–613.

Boring, Anne. 2017. "Gender Biases in Student Evaluations of Teaching." *Journal of Public Economics* 145: 27–41.

Boring, Anne, Kellie Ottoboni, and Philip B. Stark. 2016. "Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness." *ScienceOpen Research*.

Boustan, Leah, and Andrew Langan. 2019. "Variation in Women's Success across PhD Programs in Economics." *Journal of Economic Perspectives* 33 (1): 23–42.

Braga, Michela, Marco Paccagnella, and Michele Pellizzari. 2014. "Evaluating Students' Evaluations of Professors." *Economics of Education Review* 41: 71–88.

Buckles, Kasey. 2019. "Fixing the Leaky Pipeline: Strategies for Making Economics Work for Women at Every Stage." *Journal of Economic Perspectives* 33 (1): 43–60.

Carrell, Scott E., Marianne E. Page, and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *Quarterly Journal of Economics* 125 (3): 1101–44.

Carrell, Scott E., and James E. West. 2010. "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors." *Journal of Political Economy* 118: 409–32.

Dee, Thomas S. 2004. "Teachers, Race, and Student Achievement in a Randomized Experiment." *The Review of Economics and Statistics* 86 (1): 195–210.

Dee, Thomas S. 2005. "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" *American Economic Review* 95 (2): 158–65.

Deslauriers, Louis, Logan S. McCarty, Kelly Miller, Kristina Callaghan, and Greg Kestin. 2019. "Measuring Actual Learning versus Feeling of Learning in Response to Being Actively Engaged in the Classroom." *Proceedings of the National Academy of Sciences of the United States of America* 116 (39): 19251–57.

Egalite, Anna J., Brian Kisida, and Marcus A. Winters. 2015. "Representation in the Classroom: The Effect of Own-Race Teachers on Student Achievement." *Economics of Education Review* 45: 44–52.

Ehrenberg, Ronald G. 2012. "American Higher Education in Transition." *Journal of Economic Perspectives* 26 (1): 193–216.

Ehrenberg, Ronald G., and Liang Zhang. 2005. "Do Tenured and Tenure Track-Faculty Matter?" *Journal of Human Resources* 40 (3): 647–59.

Fairlie, Robert W., Florian Hoffmann, and Philip Oreopoulos. 2014. "A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom." *American Economic Review* 104 (8): 2567–91.

Feld, Jan, Nicolás Salamanca, and Ulf Zoelitz. 2020. "Are Professors Worth It? The Value-Added and Costs of Tutorial Instructors." *Journal of Human Resources* 55 (3): 836–63.

Figlio, David N., and Morton Schapiro. 2017. "Are Great Teachers Poor Scholars?" *Brookings Institution Evidence Speaks Reports* 2 (6): 1–7.

Figlio, David N., Morton O. Schapiro, and Kevin B. Soter. 2015. "Are Tenure Track Professors Better Teachers?" *The Review of Economics and Statistics* 97 (4): 715–24.

Flaherty, Colleen. 2016. "Bias against Female Instructors." *Inside Higher Education*, January 11. https://www.insidehighered.com/news/2016/01/11/new-analysis-offers-more-evidence-against-student-evaluations-teaching.

Flaherty, Colleen. 2019a. "The Dangers of Fluent Lectures." *Inside Higher Education*, September 9. https://www.insidehighered.com/news/2019/09/09/study-how-smooth-talking-professors-can-lull-students-thinking-theyve-learned-more.

Flaherty, Colleen. 2019b. "Speaking out against Student Evals." *Inside Higher Education*, September 10 https://www.insidehighered.com/news/2019/09/10/sociologists-and-more-dozen-other-professional-groups-

speak-out-against-student.

Gershenson, Seth, Cassandra Hart, Joshua Hyman, Constance Lindsay, and Nicholas Papageorge. 2018. "The Long-Run Impacts of Same-Race Teachers." NBER Working Paper 25254.

Gershenson, Seth, Stephen B. Holt, and Nicholas W. Papageorge. 2016. "Who Believes in Me? The Effect of Student-Teacher Demographic Match on Teachers Beliefs." *Economics of Education Review* 52: 209–24.

Hoffmann, Florian, and Philip Oreopoulos. 2009a. "Professor Qualities and Student Achievement." *Review of Economics and Statistics* 91 (1): 83–92.

Hoffmann, Florian, and Philip Oreopoulos. 2009b. "A Professor Like Me: The Influence of Instructor Gender on College Achievement." *Journal of Human Resources* 44 (2): 479–94.

June, Audrey Williams. 2012. "Adjuncts Build Strength in Numbers: The New Majority Generates a Shift in Academic Culture." *Chronicle of Higher Education*, November 5. https://www.chronicle.com/article/adjuncts-build-strength-in-numbers/.

Kofoed, Michael S., and Elizabeth McGovney. 2019. "The Effect of Same-Gender or Same-Race Role Models on Occupation Choice: Evidence from Randomly Assigned Mentors at West Point." *Journal of Human Resources* 54 (2): 430–67.

Lavy, Victor, and Edith Sand. 2018. "On the Origins of Gender Gaps in Human Capital: Short- and Long-Term Consequences of Teachers' Biases." *Journal of Public Economics* 167: 263–79.

Lim, Jaegeum, and Jonathan Meer. 2017. "The Impact of Teacher–Student Gender Matches." *Journal of Human Resources* 52 (4): 979–97.

Lindsay, Constance A., and Cassandra M.D. Hart. 2017. "Exposure to Same-Race Teachers and Student Disciplinary Outcomes for Black Students in North Carolina." *Educational Evaluation and Policy Analysis* 39 (3): 485–510.

Lundberg, Shelly, and Jenna Stearns. 2019. "Women in Economics: Stalled Progress." *Journal of Economic Perspectives* 33 (1): 3–22.

Mansour, Hani, Daniel I. Rees, Bryson M. Rintala, and Nathan N. Wozny. 2020. "The Effects of Professor Gender on the Post-Graduation Outcomes of Female Students." NBER Working Paper 26822.

McPherson, Michael, and Morton Owen Schapiro. 1999. "Tenure Issues in Higher Education." *Journal of Economic Perspectives* 13 (1): 85–98.

Mengel, Friederike, Jan Sauermann, and Ulf Zoelitz. 2019. "Gender Bias in Teaching Evaluations." *Journal of the European Economic Association* 17 (2): 535–66.

Morson, Gary Saul, and Morton Schapiro. 2015. *The Fabulous Future? America and the World in 2040.* Evanston, IL: Northwestern University Press.

Ran, Florence Xiaotao, and Di Xu. 2019. "Does Contractual Form Matter? The Impact of Different Types of Non-Tenure-Track Faculty on College Students' Academic Outcomes." *Journal of Human Resources* 54 (4): 1081–1120.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2014. "How Stereotypes Impair Women's Careers in Science." *Proceedings of the National Academy of Sciences* 111 (12): 4403–08.

Rivera, Lauren A., and András Tilcsik. 2019. "Scaling Down Inequality: Rating Scales, Gender Bias, and the Architecture of Evaluation." *American Sociological Review* 84 (2): 248–74.

Simonton, Teghan. 2019. "Tenure and Other Variations." *The Chronicle of Higher Education*, August 18. https://www.chronicle.com/article/tenure-and-other-variations/.

Stark, Philip, and Richard Freishtat. 2014. "An Evaluation of Course Evaluations." *ScienceOpen*.

Stroebe, Wolfgang. 2016. "Why Good Teaching Evaluations May Reward Bad Teaching: On Grade Inflation and Other Unintended Consequences of Student Evaluations." *Perspectives on Psychological Science* 11 (6): 800–816.

Supiano, Beckie. 2019. "Sociologists Caution Colleges Not to Over-Rely on Student Evaluations of Teaching." *The Chronicle of Higher Education*, September 9. https://www.chronicle.com/article/sociologists-caution-colleges-not-to-over-rely-on-student-evaluations-of-teaching/

U.S. Department of Education, National Center for Education Statistics. 2018. "Digest of Education Statistics." https://nces.ed.gov/programs/digest/d18/ (accessed December 10, 2020).

Wilson, Robin. 2010. "Tenure, RIP: What the Vanishing Status Means for the Future of Education." *Chronicle of Higher Education*, July 4. https://www.chronicle.com/article/tenure-rip-what-the-vanishing-status-means-for-the-future-of-education/.

Zafar, Basit. 2013. "College Major Choice and the Gender Gap." *Journal of Human Resources* 48 (3): 545–95.

This article has been cited by:

1. Susan Dynarski, Aizat Nurshatayeva, Lindsay C. Page, Judith Scott-Clayton. Addressing nonfinancial barriers to college access and success: Evidence and policy implications 319-403. [Crossref]
2. Наталья Владимировна Лазарева. THROUGH-THROUGH DIGITAL TECHNOLOGIES HIGHER ECONOMIC EDUCATION 79-83. [Crossref]