CellPress

Trends in Microbiology

Forum

Productivity metrics and hiring rubrics are warped by cultural schemas of merit

Mary Blair-Loy, [1,*] Stephen Reynders, [1] and Erin A. Cech [2]

Academic science is committed to objective hiring and promotion, yet diversification has stalled. Conventional approaches to improve diversity overlook the valorized cultural schemas scientists use to assess, and often mismeasure, merit. These schemas warp the design and use of productivity metrics and rubric scoring. We suggest interventions to unwarp them.

Scientists are committed to fair, objective assessments of their peers' work. Yet women and some racially minoritized academics remain under-represented in academic science, technology, engineering, and mathematics (STEM) [1]. For example, among life sciences ladder-rank faculty in 2019, 2.4% are Black, 5.0% are Latinx, and 2.8% are Black or Latinx women. The share of women (of any race) is 33.6% (authors' calculations from [1] (Table 9-29). A recent analysis of three federal datasets spanning all disciplines found that the slow rate of change in the diversity of faculty simply tracks demographic change in the American population, yet comes no closer to demographic parity [2]. Why has diversification stalled given the continued efforts to increase representation in academic science?

Conventional approaches to increasing diversity have focused on implicit bias, the cognitive shortcuts that can unfairly affect hiring and promotion decisions [3]. Yet common solutions like implicit bias training have mixed results and do not suffice as singular interventions to address systemic inequalities [4].

This article begins by explaining how preoccupation with individual-level beliefs like implicit bias overlooks another important source of the devaluation of academic scientists: the cultural frameworks STEM academics use to assess merit. These cultural schemas can distort seemingly objective measurement of merit, like citation metrics and hiring rubrics.

We introduce these schemas and then suggest interventions that departments and disciplines can use to improve the design and deployment of these measures to support more equitable hiring and promotion. Just as the problem of bias is largely rooted in scientific culture and communities, solutions also depend upon communities working together to critically assess and reconfigure their own cultural practices.

Cultural schemas of merit in academic STEM

Scholarship and policy on inequality in academic hiring and promotion often focus on implicit bias, including automatic mental associations such as linking STEM competency with men. Yet a comprehensive study of 800 organizations finds that diversity training, which often includes implicit bias training, has not produced widespread increases in employee diversity and sometimes provokes backlash [4].

A narrow focus on factors of individual cognition like implicit bias overlooks another source of inequity: widely shared cultural beliefs about STEM work. In contrast to implicit biases, which STEM faculty generally view as unfair, most faculty venerate STEM professional culture and believe that its standards are fairly applied to everyone. These shared beliefs include a set of 'cultural schemas', historically rooted, broadly shared understandings of merit, which provide moralized understandings of competence and excellence. In *Misconceiving merit*, a study of over 500 STEM faculty in an R1 university, researchers examined administrative data, a scholarly production data base, a survey, and interviews, and verified results in a broader survey of 7000 academic scientists. They discovered two cultural schemas that are core to notions of merit [5]. The schema of scientific excellence is the constellation of qualities used as indicators of merit, such as individual brilliance and assertive self-promotion. This schema tends to minimize other important criteria such as mentoring, collaboration, and diversity. Second, the work devotion schema demands single-minded allegiance to the scientific vocation and delegitimates faculty with commitments to caregiving [5].

When these two schemas are used as yardsticks to measure merit, they often mismeasure it. For example, they are more likely to identify and reward brilliance and assertive self-promotion among White and Asian heterosexual men and to overlook or even penalize these displays among Black and Latinx men [6], women of color, white women, and LGBTQ faculty. Research finds that for the same amount of scholarly productivity and net of department and job level, minoritized faculty receive less respect than their White and Asian heterosexual male colleagues and mothers receive less pay [5].

Metrics of impact

Academic hiring and promotion decisions often rely on impact metrics [7] and many scientists believe that they are largely immune to bias. Yet the two schemas can get built into these metrics in ways that perpetuate inequality.

For example, the high value put on assertive self-promotion in the scientific excellence schema smuggles acts of self-promotion into what are presumed to be measures of peer recognition. Consider the inclusion of self-citations in the H-index, which is the most recognized metric of productivity and impact [7]. Despite some debate [8,9], self-citation is generally considered an objective way of crediting one's closely related earlier research. Yet its enactment is highly gendered: men tend to self-cite much more often than women [10,11]. Nonetheless, the practice of including self-citations in individual metrics meant to measure peer recognition is seen as acceptable and nonbiased because it aligns with the scientific excellence schema's valorization of assertive self-promotion of one's scientific contributions.

Moreover, the work devotion schema mandates single-minded allegiance to a scientific calling and assumes that every year is, or should be, a research active year. This cultural belief affects the H-index, which treats all years after career launch similarly. The H-index fails to account for years with no publications, which can occur for caregivers during or following a family leave [12]. Endless working and diminished time spent caregiving is implicitly rewarded as an indicator of excellence.

Further, the scientific excellence schema venerates myths of individual brilliance and independence while overlooking the processes by which men often elevate other men in their social networks while excluding women. Given the prevalence of men in homophilic processes that exclude women [13], it is important to assess whether the practice of including self-citations in impact metrics particularly benefits male scientists. Research also finds that metrics of impact like the H-index can be manipulated by excessive and vague over-citing of former coauthors. Collaborators, often other men, may do likewise and amplify its effects [14].

Unwarping the impact metrics

We encourage associations and organizations, such as the American Association of Microbiology, the National Academies, and funders, to consider heeding the research that recommends removing self-citations from individual metrics [12] and to further investigate how reliance on impact metrics and presumptions of uninterrupted funding trajectories promote inequality. These units could recommend revisions to common metrics or propose alternative metrics. Further, they should assess how other cultural practices are hidden within metrics in ways that distort fair evaluation.

Rubrics in academic hiring decisions

To counteract the implicit bias of evaluators in hiring, diversity experts encourage the use of decision-making rubrics, in which individual evaluators use multipoint scales to rate each candidate (or semi-finalist) on a set of criteria, including factors like research productivity, impact, and teaching [15]. This process is believed to counteract individual evaluator bias by promoting more deliberative and analytical thinking and by directing attention to the skills that directly affect job performance [3].

However, two recent studies on how rubrics are used in practice find that they do not fully protect against individual bias in academic STEM. A study of rubric scoring of semi-finalist engineering faculty candidates found that women candidates received significantly lower scores on research productivity than men, even after controlling for publications, H-index, and seniority [15].

A separate study of academic faculty searches found that rubrics did little to counteract biased judgments of candidates' research productivity and impact. For example, in a life sciences department search, the belief that studies of minoritized populations were 'too narrow' led to the under-scoring and filtering out of some minoritized candidates who were conducting that research [16]. Thus, even when rubrics are used, evaluators may still define quality through the distorted lens of the schema of scientific excellence.

Unwarping the use of rubrics

These studies do not recommend abandoning rubrics. Instead, they advocate making use of rubrics in ways that utilize the emergent insights and experience of an evaluating community. One article recommends that faculty search committees meet prior to reviewing candidates' files in order to practice applying the rubrics to sample candidates and calibrate their scoring to ensure that criteria are fairly applied [16].

Another article advocates that evaluation and rubric scoring of all semi-finalist candidates should be done by all faculty members in a department, not just hiring committees [15]. It recommends that all rubric scores should be shared at the outset of the faculty meeting in order to eliminate 'first speaker bias' (when an influential colleague sets the tone by vigorously endorsing or criticizing particular candidates). And, rubric scores should be presented as one element in a holistic discussion. This included careful consideration of candidates with the largest variation in scores (who were predominantly women). Researchers found that these practices help correct misconceptions and ultimately led to greater likelihood of making job offers to women compared with the department's previous searches [15].

Concluding remarks

Implicit bias training has not ensured durable improvements in diversity [2,4]. A narrow focus on these individual biases overlooks the widely shared cultural frameworks STEM academics use to assess, and often mismeasure, merit. These schemas of merit include cultural assumptions that make the cultural conventions of impact metrics and hiring rubrics seem sensical and fair. Interventions to improve equity in hiring and promotion will be most successful if they leverage the emergent wisdom of academic communities to critically assess and reconfigure their own cultural practices.

Declaration of interests

No interests are declared.

[1]University of California, San Diego, CA, USA
[2]University of Michigan, Ann Arbor, MI, USA

*Correspondence:
mblairloy@ucsd.edu (M. Blair-Loy).

https://doi.org/10.1016/j.tim.2023.03.004

References

1. NCSES (2021) *Women, Minorities, and Persons with Disabilities in Science and Engineering*, National Center for Science and Engineering Statistics
2. Matias, J.N. *et al.* (2022) US universities are not succeeding in diversifying faculty. *Nat. Hum. Behav.* 6, 1606–1608
3. O'Meara, K. *et al.* (2020) Nudging toward diversity: applying behavioral design to faculty hiring. *Rev. Educ. Res.* 90, 311–348
4. Dobbin, F. and Kalev, A. (2022) *Getting to Diversity: What Works and What Doesn't*, The Belknap Press of Harvard University Press
5. Blair-Loy, M. and Cech, E.A. (2022) *Misconceiving Merit: Paradoxes of Excellence and Devotion in Academic Science and Engineering*, University of Chicago Press
6. Leslie, S.-J. *et al.* (2015) Expectations of brilliance underlie gender distributions across academic disciplines. *Science* 347, 262–265
7. Simoes, N. and Crespo, N. (2020) Self-citations and scientific evaluation: leadership, influence, and performance. *J. Informetr.* 14, 1–13
8. Szomszor, M. *et al.* (2020) How much is too much? The difference between research influence and self-citation excess. *Scientometrics* 123, 1119–1147
9. Van Noorden, R. and Chawla, D.S. (2019) Policing self-citations. *Nature* 527, 578–579
10. King, M.M. *et al.* (2017) Men set their own cites high: gender and self-citation across fields and over time. *Socius* 3, 1–22
11. Dworkin, J.D. *et al.* (2020) The extent and drivers of gender imbalance in neuroscience reference lists. *Nat. Neurosci.* 23, 918–926
12. Cameron, E.Z. *et al.* (2016) Solving the productivity and impact puzzle: do men outperform women, or are metrics biased? *BioScience* 66, 245–252
13. Cyr, E.N. *et al.* (2021) Mapping social exclusion in STEM to men's implicit bias and women's career costs. *Proc. Natl. Acad. Sci. U. S. A.* 118, 1–7
14. Poirrier, M. *et al.* (2021) Robust h-index. *Scientometrics* 126, 1969–1981
15. Blair-Loy, M. *et al.* (2022) Can rubrics combat gender bias in faculty hiring? *Science* 377, 35–37
16. White-Lewis, D.K. (2020) The facade of fit in faculty search processes. *J. High. Educ.* 91, 833–857