Board 430: What Constitutes Research Excellence? Experimental Findings on Factors Driving Faculty Perceptions of Tenure Candidates in STEM

Dr. John K. Wagner, University of New Mexico

Dr. John K. Wagner is a Postdoctoral Fellow for the Alliances for Graduate Education and the Professoriate (AGEP) joint project between the University of New Mexico, Arizona State University, and the University of Oregon. His research interests focus on improving equity in higher education, and the impacts of technology on public opinion and political psychology. Specifically, Dr. Wagner's research explores how to improve equity and inclusion by measuring latent support for exclusionary practices, and designing interventions targeted at improving gaps in graduate student success and support, and faculty hiring, tenure, and promotion. Additionally, his research attempts to understand the impacts of online platforms and their algorithms on political contention in the United States.

Lizandra C. Godwin, University of New Mexico

Dr. Lizandra C. Godwin is a faculty member in the Electrical and Computer Engineering Department at the University of New Mexico. Dr. Godwin earned her M.S. and Ph.D. degrees in Materials Science and Engineering from the University of Florida, and her B.S. in Mechanical Engineering from Florida A&M University. Her research focuses on materials (ink) development for advanced manufacturing processes, microelectronic devices, and broadening participation in engineering through her engineering education research. Prior to academia, Dr. Godwin worked for many companies, including Intel Corp.

Magda Hinojosa, Arizona State University

Dr. Magda Hinojosa is Professor of Political Science and Dean of Social Sciences in the College of Liberal Arts and Sciences at Arizona State University. She received her B.A. in government from the University of Texas at Austin and her A.M. and Ph.D. in political science from Harvard University. Dr. Hinojosa's work examines women's political incorporation in Latin America, including the role of candidate selection procedures in explaining women's underrepresentation in formal positions of power and the effects of women's political representation. Her work has been supported by the Ford Foundation, Fulbright, the National Science Foundation, and USAID. She is the author of Selecting Women, Electing Women: Political Representation and Candidate Selection in Latin America (Temple University Press, 2012) and co-author of Seeing Women, Strengthening Democracy: How Women in Politics Foster Connected Citizens (Oxford University Press, 2020).

Dr. Nadia Singh, University of Oregon

Dr. Nadia Singh is an Associate Professor of Biology at the University of Oregon. Dr. Singh earned her B.A. magna cum laude from Harvard University, earned her PhD in Biological Sciences from Stanford University, and did a postdoctoral fellowship at Cornell University. She was faculty at North Carolina State University for seven years before joining the University of Oregon. Dr. Singh's research focuses on evolutionary genetics and uses Drosophila melanogaster as a model system.

Elizabeth A. Wentz, Arizona State University

Dr. Elizabeth A. Wentz is the Vice Provost and Dean of the Graduate College and Professor in the School of Geographical Sciences and Urban Planning at Arizona State University. She earned her PhD in Geography from the Pennsylvania State University and MA and BS from the Ohio State University. Her research focuses on geospatial technologies in urban adaptation to climate change.

Benjamin Jose Aleman, University of Oregon
Prof. Mala Htun, University of New Mexico

Mala Htun is Professor of Political Science at the University of New Mexico and Special Advisor for Inclusion and Climate in the School of Engineering. She works on women's rights, social inequalities, and

strategies to promote inclusion and diversity. Htun is the author of three books, most recently The Logics of Gender Justice: State Action on Women's Rights around the World, co-authored with Laurel Weldon (Cambridge Press, 2018). Htun is lead PI of the NSF-AGEP Catalyst Alliance of UNM, ASU, and U-Oregon and PI of the NSF-funded New Strategies to Combat Harassment in Engineering research project. She has been an Andrew Carnegie Fellow, a fellow at the Kellogg Institute of the University of Notre Dame and the Radcliffe Institute of Harvard, and held the Council on Foreign Relations International Affairs Fellowship in Japan. She holds a Ph.D. in political science from Harvard and a A.B. in international relations from Stanford. She was an assistant and then associate professor at the New School for Social Research from 2000-2011.

What Constitutes Research Excellence? Experimental Findings on Factors Driving Faculty Perceptions of Tenure Candidates in STEM

Introduction

Hiring, academic reviews, and tenure and promotion (T&P) are the most important checkpoints along the academic career path in STEM. The hiring process shapes the sub-field and demographic composition of academia, while annual reviews dictate advancement to promotion, awards, and salary. Tenure is a particularly high-stakes juncture, as it sets up a decades-long relationship with faculty colleagues, and grants life-long job security while also conferring a badge of honor and legitimacy in the global scientific community. What determines whether or not a scholar passes through these critical academic checkpoints?

Evaluation by faculty colleagues—both inside and outside the institution—determines hiring and promotion outcomes. While teaching and service are typical components of evaluation, the most important component of a candidate's dossier is their research record [1]–[4]. A record which also greatly influences the ability of an individual to secure external private/ public funding. The research record typically consists of a mixture of quantitative information (such as external grant revenue, number of publications or awards, *h*-index, etc.), categorical variables (such as the prestige of PhD and postdoctoral training institution), and qualitative assessments about the quality and novelty of research.

There are two fundamental challenges associated with the evaluative component of these academic checkpoints. In the first place, many academic units lack clear standards to assess candidates and most faculty have different understandings of the relative importance of different evaluation criteria. Committees therefore have considerable discretion to evaluate different candidates in different ways. The lack of consistency makes the T&P process rife with potential for bias [5]–[11] and susceptible to power politics [12], [13].

The second challenge associated with evaluation is that the indicators of "research excellence" that are often employed are inherently biased against women and people of color, and even their research focus [14], [15]. When it comes to external funding, non-white scientists are less likely to receive NSF grants [16] and NIH funding rates for African-American scientists are lower than for whites [17], [18]. With regard to the citation counts informing the h-index, women receive fewer citations than men [19], [20]. Journal impact factors may also screen out underrepresented groups, since women and minority scholars and scientists are underrepresented among authors in the most highly-ranked journals [21], [22].

The first step in creating new policies and procedures for equitable evaluation at each of these critical checkpoints is to understand the nature of the evaluation process. *What factors and factor interactions determine research excellence? How does the relative importance of different evaluation criteria vary across individuals, fields, and institutions?*

To address these questions, we designed an experiment to unpack research excellence. We created fictitious tenure and promotion candidates with different combinations of key performance metrics (e.g. number of publications, h-index, etc.) and asked STEM faculty to evaluate them holistically. Our data show that external grant funding, number of publications, and h-index were the most significant contributors to evaluation of overall research excellence. Interestingly, the prestige of PhD-granting institution was not a significant contributor overall.

Our findings reveal that a seemingly quantitative evaluation process is subjective and variable and may therefore reproduce, and not correct for, societal biases and inequalities. Our study sheds light on the obstacles to the career advancement of scientists and scholars who identify as members of underrepresented groups [23], [24]. In addition, although we studied a targeted set of factors and focus on STEM, the tools we developed are valuable to examine evaluation processes with different sets of criteria, across disciplines, at multiple career-stages.

In this paper, we address the first of our research questions about the determinants of research excellence. We present aggregate results that take into account all the participants in our study. In future works, we address the second question about variation by analyzing differences across individuals, fields of study, and institution.

Methods

To answer measure and model the concept of "research excellence," we employed a factorial experiment, packaged as an online survey, to STEM faculty members at 3 public, R1 institutions. The goal of the experiment was to identify the relative importance of the multiple criteria that gatekeepers and evaluators normally use to assess the research achievements of candidates for T&P.

The factorial experiment. We designed a blocked seven-factor, two-level fractional factorial experiment and measured one response variable–research excellence, shown in Figure 1. We targeted seven factors to test for contributions to research excellence through the lens of promotion and tenure. The factors (i.e. predictor variables) and corresponding ± levels (parenthesis) include: h-index (5, 25), prestige of PhD institution (average, high), external funding (intramural, NSF CAREER/NIH R01), number of publications since appointed (10, 30), whether the candidate published in high-prestige journals (No, Yes), and the desirability of the research focus (mainstream to discipline, outside the disciplinary box), and view (your colleague's view, your view). We set all factors to one of the two discrete ± levels. The experimental response (i.e. dependent variable) is "research excellence" measured on a scale of $1 - 5$, with 5 as "most excellent."

We used a 5+ resolution design with 64 unique treatments (i.e. a 26 fractional) and a balanced distribution of factor levels for each factor (i.e. each ± factor level is represented 32 times). The design resolution enables a simultaneous model with main effects, all two-factor interactions, and a mixture of aliased three-factor interactions. We employed eight-run blocks to filter evaluator-to-evaluator scale variation and randomized treatments within each block.

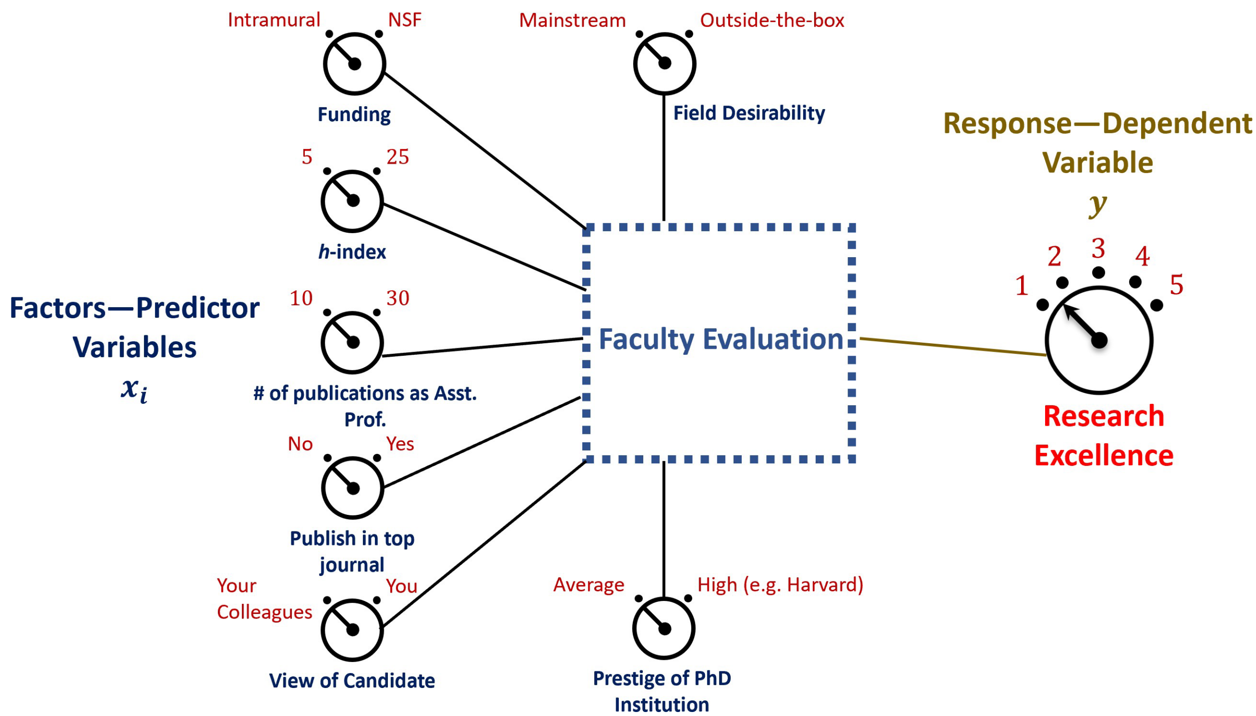

Figure 1: Factorial experiment design for determining the relative impact of research metrics on faculty-reported research excellence. Potential predictors are displayed as options for the research participant- acting as a faculty evaluator. Note that all predictors are binary, and research excellence is captured on a 1-5 scale.

The factorial survey instrument. The survey presented STEM faculty participants with summarized research profiles of different hypothetical candidates going up for tenure and asked them to score the candidates' research excellence. Figure 2 shows an example of a candidate profile. The record table lists six of the factors and their values, while the seventh "view" factor is built into the evaluation prompt. The prompt reads: "Please indicate how [ ] would evaluate this candidate" with either "your colleagues" or "you" in the prompt [ ].

Each candidate profile represents one of the 64 experimental treatments. Experimental blocking required each participant to review eight different, randomized candidates of the 64 candidate profiles. Once observations for all 64 profiles were obtained, the order of profiles was randomized again and the next set of respondents viewed eight of the 64 candidate profiles, albeit in a new randomized order.

Candidate's Research Record:

h-index	5
Ph.D. Institution	Average prestige R1, e.g. Montana State University, Univ of Georgia
Research Funding	NSF CAREER or NIH R01
Published in Top Journal?	No
Number of publications since hire as junior faculty	10
Research Topic	Outside the disciplinary box

Please indicate **how your colleagues** would evaluate this candidate, with 5 being the most excellent, and 1 being the least excellent.

1	2	3	4	5

Figure 2: Example profile of the tenure-promotion file for participant's review. Each profile appeared as one of eight files on two pages of a Qualtrics survey. The values in the right- hand column of the table and the underlined portion of the prompt text were altered as the factorial design demanded.

The survey was deployed through Qualtrics to all STEM tenured/ tenure-track faculty at three R1 universities, which represent different geographical regions, the full spectrum of active STEM disciplines, and a gradient of underrepresented minority representation. The survey offered participants compensation for their time. The team invited 516 faculty to participate, 174 responded to the survey invitation, and 170 completed it. The resultant response rate and completion rate are 33.7% and 98% respectively. The survey was initially deployed November 9th, 2022, with some departments receiving distributions later between November 14th and 18th. All contacts were sent two reminders during the data collection window. The survey was closed on November 30th. In total, observations made for the whole set of 64 candidates were replicated 19 times for a total of 1216 runs and 164 distinct respondents.

In addition, we presented participants with a set of multiple-choice survey questions about factors affecting tenure and promotion decisions. Surveying respondents *in addition to* the factorial experiment enabled the team to assess how respondents perceive their preferences for certain metrics of research excellence versus how they actually evaluate candidates holistically as shown in the experiment. The team's ability to point to any discrepancies between perceived and exhibited evaluation behaviors will help show actual faculty evaluators how their perceived evaluations do or do not match observed evaluation behaviors. The battery presented to respondents consisted of twelve questions (listed below) asking respondents how important each factor is to them *personally* when assessing the research excellence of a tenure-promotion file. Each question was evaluated on a six-point Likert scale ranging from (1) "Not at all important" to (6) "Very Important."

- Assessments provided by external reviewers.
- Whether the candidate has received significant federal funding, such as a NSF Career Award or NIH R01.
- The candidate's number of publications.
- Assessment of colleagues in the candidate's academic unit.
- Whether the candidate is conducting research that is novel, bold and outside the disciplinary box.
- The placement of publications in top journals of the field with high impact factors.
- Whether the candidate uses research methods endorsed by the mainstream of the discipline and regularly published in high impact journals.
- The candidate's h-index.
- Whether the candidate is conducting research on topics that are central to the disciplinary mainstream.
- Whether the candidate's research has clear practical applications and real-world relevance.
- Prestige of the institution where candidate held a post-doctoral fellowship.
- Prestige of the institution where candidate received their Ph.D.

Data analysis

The factorial experiment data were modeled with a categorical effects model that included up to two-factor interactions,

$$\hat{y} = \beta_0 + \sum_{i=1}^{7} \beta_i x_i + \sum_{i,j,i\neq j}^{7} \beta_{i,j} x_i\, x_j$$

where $\hat{y}$ is the predicted research excellence, β_0 is the intercept (i.e., mean measured research excellence), β_i is the main effect coefficient for the i^th factor x_i, and $\beta_{i,j}$ is the interaction coefficient for factors x_i and x_j. The model was reduced by using an analysis of variance (ANOVA) framework (JMP software, version 16) to remove statistically insignificant effects and interactions with an $\alpha = 0.05$ cutoff. We retained insignificant coefficients only when needed to preserve model hierarchy. The blocking variable is retained with α =0.05. We then verified the ANOVA modeling assumptions, including normal, independently distributed residuals, and equality of variance. This framework is then used to create a quantitative, predictive model for the response ($\hat{y}$) in terms of the input factors (x_i). The analysis quantifies the relative strength of each factor (main effects) and factor interactions (higher-order effects), and determines which factors have no significant effect on the response.

Findings

The results of the factorial experiment are displayed in Figure 3, which displays the β_i coefficients and p-value for the intercept, all main effects, and the single two-way interaction which displayed significance ($p < 0.05$). Four singular factors exhibit a significant effect on research excellence evaluations.

External funding presents by far the largest effect, resulting in a significant leap in evaluations of research excellence ($p < 0.0001$). A candidate with a record of winning federal funding such as a NSF Career Grant, when compared to intramural funding, increases estimated evaluations of tenure files by 0.9 points.

A higher h-index presents the second largest effect, leading to a significant leap in evaluations of research excellence ($p < 0.0001$). An h-index of 25 as opposed to 5 increases estimated evaluations of tenure files by 0.54 points. Third, a greater number of publications results in a significant leap in evaluations of research excellence ($p < 0.0001$). 30 publications after hiring to the tenure track position as opposed to 5 increases estimated evaluations of tenure files by 0.42 points. Finally, publishing in a prestigious, high-impact journal results in a significant leap in evaluations of research excellence ($p < 0.0001$) and increase estimated evaluations of tenure files by 0.42 points.

Notably, Ph.D. institution prestige does not significantly affect evaluations of research excellence. Desirability of a researcher's field ($p = 0.10$) and whether the evaluation was "yours" or "your colleague's" does not significantly affect ($p = 0.62$) evaluations of research excellence alone. However, the interaction of a candidate's field desirability with the "View of candidate" factor is significant ($p = 0.017$). This interaction indicates that the change in the evaluation score when the candidate's discipline is switched from "mainstream" to "outside the mainstream" depends on whether the evaluation is "yours" or "your colleagues". Here, the score increases when the evaluator grades the candidate and decreases when the evaluators' colleagues grade. One interpretation of this result is that evaluators feel they value outside- the-box research more than their colleagues.

Descriptive results of the survey questions are displayed in Figure 4. The figure orders the questions according to the proportion of respondents who name a factor that they themselves believe to be at least somewhat important when evaluating research excellence of a tenure-promotion file. For example, 99% of respondents report that assessments by external reviewers are important. Given such external reviewers often review files based on characteristics similar to those displayed in the factorial experiment above, the experimental results also are relevant as they impact internal reviewers via external reviewer's assessments.

i	x_i Factors	β_i Model coefficient	p-value
0	*Intercept*	3.72	< 0.0001
1	**External Funding**	0.45	< 0.0001
2	***h*-index**	0.27	< 0.0001
3	**# of pubs**	0.21	< 0.0001
4	**Publishing prestige**	0.21	< 0.0001
5	Field desirability	-0.01	0.097
6	PhD prestige	0.02	0.252
7	View of candidate	-0.01	0.623
5,7	**View of candidate × Field desirability**	0.05	0.017

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_{57} x_5 x_7$$

Figure 3: Effects model from pilot experiment. $\hat{y}$ is the research excellence score predicted by the preliminary data. The β_i coefficients are calculated for factors x_i with coded levels (±1).

Secondly, and similar to the relative ordering of the main effects in the factorial results, federal funding is considered second most important on average. Next, and largely mirroring the order of the factorial results, are the candidate's number of publications, the assessment of colleagues in a candidate's academic unit, outside the box research, publications in top journals, the use of mainstream methods, a candidate's h-index, and mainstream research topics. It is notable that the practical applicability of the candidate's research, as well as the prestige of Post-doc and Ph.D. institutions, are considered unimportant by a majority (> 50%) of the respondents.

While perceptions of evaluating the research excellence of a tenure-promotion file in the survey largely conforms to the evaluation behaviors displayed in the factorial experiment, there are some key differences. One notable distinction between the survey and factorial results is that while the h-index is the second most influential factor in the experiment, it is ranked in the lower half of factors in the survey behind the number of publications and publishing in top journals.

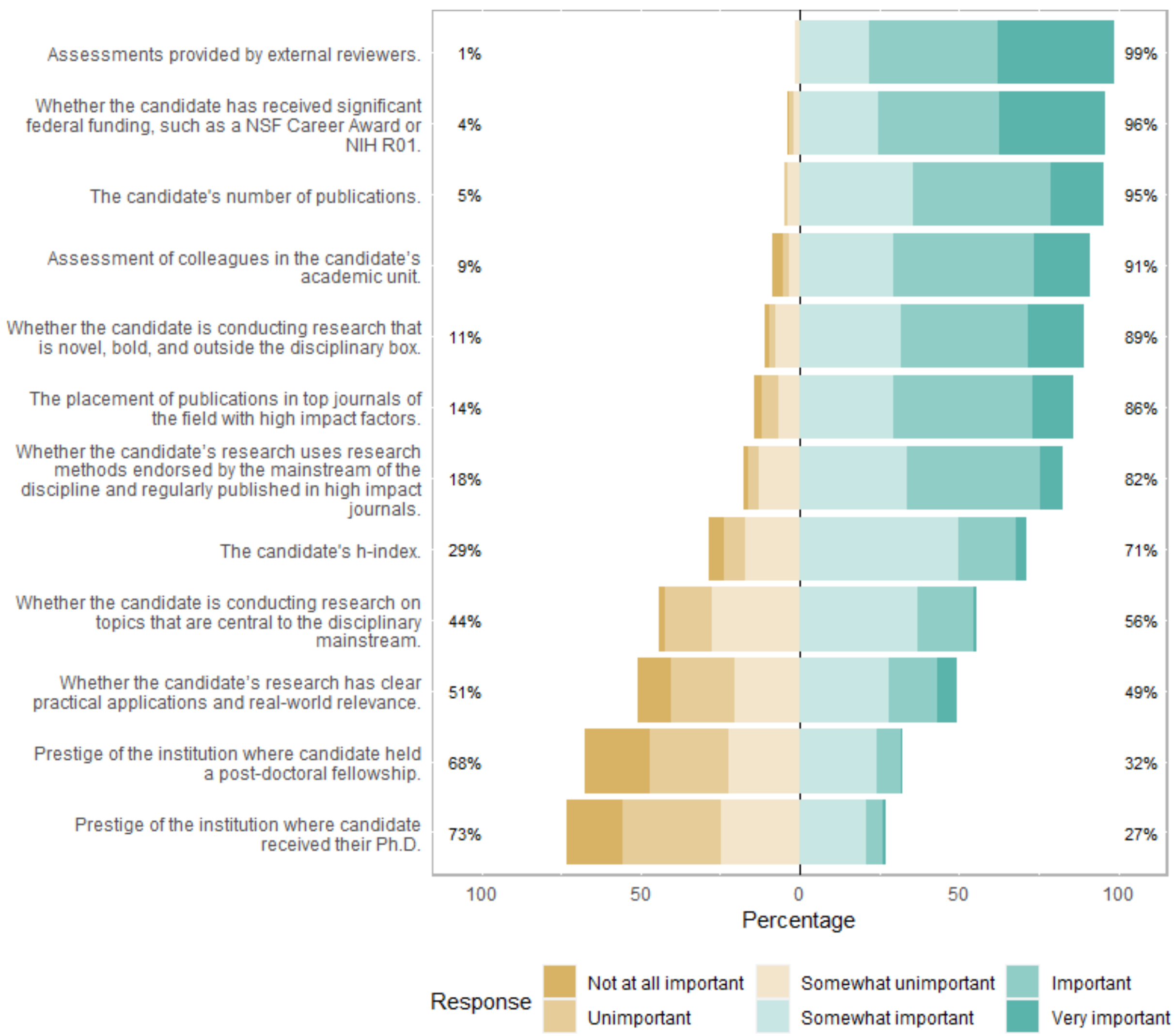

Figure 4: Likert plot of the reported importance of aspects of tenure-promotion candidate's research excellence. The left-hand percentages represent the percentage of respondents naming an aspect as unimportant, and the right-hand percent important. Aspects are ordered top-to-bottom by greatest percentage considered important.

Discussion

The findings from our experiment and survey emphasize the importance of external research funding, followed by h-index and number of publications, as measurements of research excellence in current practice. Previous studies have presented evidence of the importance of these criteria to faculty (for a review, see [2]) yet ours may be the first study to quantify the impact of each of these factors using experimental methods.

Our study finds that external funding is the single most important criterion influencing perceptions of research excellence. Controlling for other factors, candidates with a history of receiving a NSF Career Award or a NIH R01 grant receive a score that is 20% higher than candidates with a record of only intramural funding. Other factors matter too: number of publications and h-indexes are also significant impacts on research excellence scores.

It is notable that the prestige of the candidate's PhD institution is not a significant factor in our analysis, since in large, observational studies, institutional prestige is the major driver of placement, productivity, and success winning grants [25]–[27]. This suggests that institutional prestige may be a confounding variable that the external funding factor implicitly measures in our factorial analysis.

Conclusion

Many universities seek to diversify their faculty ranks, particularly in STEM fields where African American, Hispanic-Latino, and Native American scientists are few in number. Academic institutions value diversity to advance equality and justice for historically marginalized groups. There is also evidence that socially diverse research enterprises produce more influential, relevant, and better results [28]–[30].

This paper presents a fresh analysis on the challenge of faculty diversity by examining the tenure and promotion process. We find that, overall, external funding, h-index, and publication count are the most important factors driving faculty perceptions of research excellence sheds light on some of the barriers to diversifying STEM faculty ranks. We know that these metrics put scholars from underrepresented groups at a disadvantage. To increase diversity among faculty, we need to develop strategies to direct external funding toward underrepresented groups, compel committees to consider additional criteria of excellence, and emphasize qualitative deliberation, rather than quantitative scoring, in faculty evaluation.

Acknowledgements

This material is based upon work supported by the National Science Foundation (NSF) under Grant No. 2149204, Alliances for Graduate Education and the Professoriate Catalyst Alliance (AGEP CA). This AGEP CA promotes systemic change by investigating how the T&P evaluation practice might adversely affect scholars who identify as African American, Hispanic American, American Indian, Alaska Native, Native Hawaiian, and/ or Native Pacific Islander. Our goal is to contribute to scientific knowledge on the strategies to improve the presence and success of AGEP populations in the STEM professoriate. The research discussed in this paper in one of the initial steps we have taken towards that goal.

References

[1] J. M. Carey, K. R. Carman, K. P. Clayton, Y. Horiuchi, M. Htun, and B. Ortiz, "Who wants to hire a more diverse faculty? A conjoint analysis of faculty and student preferences for gender and racial/ethnic diversity," *Politics, Groups, and Identities*, vol. 8, no. 3, pp. 535–553, May 2020, doi: 10.1080/21565503.2018.1491866.

[2] L. A. Schimanski and J. P. Alperin, "The evaluation of scholarship in academic promotion and tenure processes: Past, present, and future," *F1000Res*, vol. 7, p. 1605, Oct. 2018, doi: 10.12688/f1000research.16493.1.

[3] D. B. Rice, H. Raffoul, J. P. A. Ioannidis, and D. Moher, "Academic criteria for promotion and tenure in biomedical sciences faculties: cross sectional analysis of international sample of universities," *BMJ*, vol. 369, p. m2081, Jun. 2020, doi: 10.1136/bmj.m2081.

[4] N. A. Bonn and W. Pinxten, "Advancing science or advancing careers? Researchers' opinions on success indicators," *PLOS ONE*, vol. 16, no. 2, p. e0243664, Feb. 2021, doi: 10.1371/journal.pone.0243664.

[5] S. W. Davies *et al.*, "Promoting inclusive metrics of success and impact to dismantle a discriminatory reward system in science," *PLOS Biology*, vol. 19, no. 6, p. e3001282, Jun. 2021, doi: 10.1371/journal.pbio.3001282.

[6] B. Mitchneck, J. L. Smith, and M. Latimer, "A recipe for change: Creating a more inclusive academy," *Science*, vol. 352, no. 6282, pp. 148–149, Apr. 2016, doi: 10.1126/science.aad8493.

[7] A. Clauset, S. Arbesman, and D. B. Larremore, "Systematic inequality and hierarchy in faculty hiring networks," *Science Advances*, vol. 1, no. 1, p. e1400005, Feb. 2015, doi: 10.1126/sciadv.1400005.

[8] F. R. Jensenius, M. Htun, D. J. Samuels, D. A. Singer, A. Lawrence, and M. Chwe, "The benefits and pitfalls of Google Scholar," *PS: Political Science & Politics*, vol. 51, no. 4, pp. 820–824, 2018.

[9] P. E. Smaldino and R. McElreath, "The natural selection of bad science," *Royal Society Open Science*, vol. 3, no. 9, p. 160384, Sep. 2016, doi: 10.1098/rsos.160384.

[10] J. J. Heckman and S. Moktan, "Publishing and Promotion in Economics: The Tyranny of the Top Five," *Journal of Economic Literature*, vol. 58, no. 2, pp. 419–470, Jun. 2020, doi: 10.1257/jel.20191574.

[11] M. D. Lindner, K. D. Torralba, and N. A. Khan, "Scientific productivity: An exploratory study of metrics and incentives," *PLOS ONE*, vol. 13, no. 4, p. e0195321, Apr. 2018, doi: 10.1371/journal.pone.0195321.

[12] E. R. Auster and A. Prasad, "Why Do Women Still Not Make It to the Top? Dominant Organizational Ideologies and Biases by Promotion Committees Limit Opportunities to Destination Positions," *Sex Roles*, vol. 75, no. 5, pp. 177–196, Sep. 2016, doi: 10.1007/s11199-016-0607-0.

[13] L. A. Rivera, "When two bodies are (not) a problem: Gender and relationship status discrimination in academic hiring," *American Sociological Review*, vol. 82, no. 6, pp. 1111–1138, 2017.

[14] I. H. Settles, L. R. Warner, N. T. Buchanan, and M. K. Jones, "Understanding psychology's resistance to intersectionality theory using a framework of epistemic exclusion and invisibility," *Journal of Social Issues*, vol. 76, no. 4, pp. 796–813, 2020, doi: 10.1111/josi.12403.

[15] K. Dotson, "Conceptualizing Epistemic Oppression," *Social Epistemology*, vol. 28, no. 2, pp. 115–138, Apr. 2014, doi: 10.1080/02691728.2013.782585.

[16] C. Y. Chen *et al.*, "Systemic racial disparities in funding rates at the National Science Foundation," *eLife*, vol. 11, p. e83071, Nov. 2022, doi: 10.7554/eLife.83071.

[17] T. A. Hoppe *et al.*, “Topic choice contributes to the lower rate of NIH awards to African-American/black scientists,” *Science Advances*, vol. 5, no. 10, p. eaaw7238, Oct. 2019, doi: 10.1126/sciadv.aaw7238.
[18] D. K. Ginther *et al.*, “Race, Ethnicity, and NIH Research Awards,” *Science*, vol. 333, no. 6045, pp. 1015–1019, Aug. 2011, doi: 10.1126/science.1196783.
[19] G. Ghiasi, V. Larivière, and C. R. Sugimoto, “On the Compliance of Women Engineers with a Gendered Scientific System,” *PLOS ONE*, vol. 10, no. 12, p. e0145931, Dec. 2015, doi: 10.1371/journal.pone.0145931.
[20] D. Maliniak, R. Powers, and B. F. Walter, “The Gender Citation Gap in International Relations,” *International Organization*, vol. 67, no. 4, pp. 889–922, Oct. 2013, doi: 10.1017/S0020818313000209.
[21] D. L. Teele and K. Thelen, “Gender in the Journals: Publication Patterns in Political Science,” *PS: Political Science & Politics*, vol. 50, no. 2, pp. 433–447, Apr. 2017, doi: 10.1017/S1049096516002985.
[22] P. Chakravartty, R. Kuo, V. Grubbs, and C. McIlwain, “#CommunicationSoWhite,” *Journal of Communication*, vol. 68, no. 2, pp. 254–266, Apr. 2018, doi: 10.1093/joc/jqy003.
[23] L. Urrieta, L. Méndez, and E. Rodríguez, “‘A moving target’: a critical race analysis of Latina/o faculty experiences, perspectives, and reflections on the tenure and promotion process,” *International Journal of Qualitative Studies in Education*, vol. 28, no. 10, pp. 1149–1168, Nov. 2015, doi: 10.1080/09518398.2014.974715.
[24] A. A. Berhe *et al.*, “Scientists from historically excluded groups face a hostile obstacle course,” *Nat. Geosci.*, vol. 15, no. 1, Art. no. 1, Jan. 2022, doi: 10.1038/s41561-021-00868-0.
[25] K. H. Wapman, S. Zhang, A. Clauset, and D. B. Larremore, “Quantifying hierarchy and dynamics in US faculty hiring and retention,” *Nature*, vol. 610, no. 7930, pp. 120–127, 2022.
[26] W. P. Wahls, “High cost of bias: Diminishing marginal returns on NIH grant funding to institutions.” bioRxiv, p. 367847, Jul. 13, 2018. doi: 10.1101/367847.
[27] M. A. Taffe and N. W. Gilpin, “Racial inequity in grant funding from the US National Institutes of Health,” *eLife*, vol. 10, p. e65697, Jan. 2021, doi: 10.7554/eLife.65697.
[28] L. Smith-Doerr, S. N. Alegria, and T. Sacco, “How Diversity Matters in the US Science and Engineering Workforce: A Critical Review Considering Integration in Teams, Fields, and Organizational Contexts,” *Engaging Science, Technology, and Society*, vol. 3, pp. 139–153, Apr. 2017, doi: 10.17351/ests2017.142.
[29] M. W. Nielsen *et al.*, “Gender diversity leads to better science,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 8, pp. 1740–1742, Feb. 2017, doi: 10.1073/pnas.1700616114.
[30] B. K. AlShebli, T. Rahwan, and W. L. Woon, “The preeminence of ethnic diversity in scientific collaboration,” *Nat Commun*, vol. 9, no. 1, Art. no. 1, Dec. 2018, doi: 10.1038/s41467-018-07634-8.