School Leadership Review

Volume 20
Issue 1 *Spring/Summer 2025 Continuing A Legacy of Leadership*

Article 7

March 2025

Dismantling inequities in the faculty evaluation system

Mary Petron
Sam Houston State University, map042@shsu.edu

William Blackwell
Sam Houston State University

Abbie Strunc
Sam Houston State University

DISMANTLING INEQUITIES IN THE FACULTY EVALUATION SYSTEM

Abstract

Qualifying for merit is a competitive and contentious process in higher education. Many merit structures are based upon traditional faculty evaluation systems which do not reflect the changing nature of higher education and thus, yield inequitable results. This article presents the case of a large department at a mid-size regional university that revised the system for merit and faculty evaluation to recognize undergraduate contributions and more equitably score the scholarship, service, and teaching of faculty members in the department. The case describes the relevant literature, the university and departmental context, the process for designing the new evaluation system, and lessons learned from the process. This case illustrates the challenges and possibilities associated with developing merit-based faculty evaluation systems that incentivize faculty productivity and provide a meaningful and fair assessment of faculty performance.

Keywords: *faculty evaluation, merit pay, higher education*

Dismantling Inequities in the Faculty Evaluation System

Introduction

Merit pay is a highly contentious issue for faculty at many institutions of higher education, especially when performance evaluation is directly connected to merit awards and the amount granted (Buller, 2009). In Texas, public colleges and universities were mandated by the Texas legislature to create and implement a system of evaluation and merit pay for performance as the sole basis for salary increases for faculty, which resulted in the end of cost of living pay increases (Hunnicutt, et al., 1991). Supporters of merit systems contend these systems provide an incentive and reward for professors who make exemplary contributions to the institution's mission (Buller, 2009). They argue that linking pay increases to performance will contribute to increased productivity (Miller, 1988; Buller, 2009; Field, 2015).

Opponents of merit systems assert that it is disincentivizing for faculty due to the small availability of funds found in higher education merit pools (Buller, 2009). Faculty members who are top performers question whether a 4% raise is worth the effort and low performers tend to produce less because even their limited effort is not recognized (Buller, 2009). Additionally, faculty pay is not that varied among departments within a college, but rather the difference is stark between colleges. For example, professors in colleges of business and law make more than those in the humanities and social sciences (Bichsel et al., 2021).

This article presents the case study of a large department at a mid-size regional university in Texas that recently revised the system for faculty evaluation and merit to recognize undergraduate contributions and more equitably score the scholarship, service, and teaching of faculty members in the department. The case describes the relevant literature, the university and

departmental context, the process for designing the new evaluation system, and lessons learned from the process. This case study illustrates the challenges and possibilities associated with developing merit-based faculty evaluation systems that incentivize faculty productivity and provide a meaningful and objective assessment of faculty performance.

Literature Review

Faculty evaluation

Faculty performance in higher education, and the ways in which merit-based systems are designed and implemented, have been examined by researchers since the latter half of the twentieth century. Field (2015) found in her review of the literature that studies on merit-based faculty evaluation systems fell into three categories: a) merit pay is warranted in principle and works well in practice; b) merit pay is warranted in principle, but is a failure in practice, and c) merit pay fails in both principle and practice. Concerns raised related to merit-based faculty evaluation systems have found that these systems are inequitable in practice and have a negative impact on career trajectory of many faculty members (Laube et al., 2007; Williams, 2007). Research into inequities within evaluation systems have examined different structures between colleges and universities, gender, ethnicity, student performance, student perception, and recently, the impact of the pandemic (Gannon, 2021) across the three main areas of evaluation: scholarship, teaching and service. Each of these three areas is discussed below.

Scholarship

Research has identified multiple challenges related to the evaluation of faculty scholarship. One area of challenge relates to limitations in the metrics used to evaluate scholarly productivity (Schimanski & Alperin, 2018). Traditional metrics, such as journal reputation and impact factor, privilege publication in a narrow range of outlets within more popular or

established fields of research. These metrics favor researchers working in more established fields or faculty who have research agendas that align with the focus of longer-standing journals. As a result, the recognition given to scholarly productivity in specialized or emerging fields, or those fields that focus on minoritized populations, is minimized in many faculty evaluation systems (Schimanski & Alperin, 2018; Settles et al., 2021).

This challenge directly connects to discrepancies in the evaluation of women and faculty of color. A significant area of research into faculty performance has been the perceived and verified discrepancies for women and faculty of color. Settles et al. (2021) conducted a study in a predominantly white, research-intensive university. The findings revealed that most of the 118 faculty of color interviewed felt their scholarship was consistently devalued and had to be proven to meet university standards more than the work of their colleagues. The faculty reported that the formal processes for evaluation "worked operationally by privileging certain journals or publishing presses as outlets for quality work" while viewing publications not within the privileged group as lesser quality work (Settles et al., 2021, p. 502).

A final challenge related to the evaluation of scholarship relates to the teaching responsibilities of faculty. Faculty in service-oriented fields such as teaching and nursing often have heavy teaching loads with significant time committed to supervising candidates in the field. Although these faculty members have positions that require them to prioritize teaching, their evaluation systems do not take this into account (Schimanski & Alperin, 2018). Instead, scholarly productivity via the number of articles in high profile journals and ability to secure external grant funding is still prioritized over their teaching contributions. Evaluation systems have not effectively defined and measured scholarly productivity for faculty in more teaching-focused roles who do not have the time or opportunity to engage in scholarship that conforms to

traditional conceptions of scholarly productivity in research universities (Smith & Walker, 2022).

Teaching

Research has revealed barriers associated with the evaluation of faculty teaching effectiveness. A primary challenge relates to the use of student evaluations of their professors. This is a particularly impactful area of evaluation, as student evaluations of faculty teaching are frequently primary considerations in determining potential merit raises, as well as promotion, tenure, and retention decisions. The literature has identified inherent biases in student evaluations of teaching and suggests that students are unlikely to be competent evaluators of good teaching (Hunnicutt et al., 1991; Arugete et al., 2017; Gannon, 2021). Reports indicate gender, race, instructional discipline, demographic characteristics, and perceived attractiveness of the professor influence student evaluations of faculty teaching (Arugete et al., 2017). Women and faculty of color are negatively affected by these biases. Research from Akins and Murphy (2019) and Mitchell and Martin (2018) highlights these discrepancies and demonstrates the need for teaching evaluation systems that provide balanced approaches that do not over-emphasize student evaluations.

The difficulties with student evaluations of teaching further exacerbate other struggles with assessing teaching quality. Concerns have been raised related to the ways in which department chairs evaluate faculty teaching performance and a lack of structured teaching evaluation instruments that can be used for peer-evaluation (Drue & Bifulco, 2025). Additionally, the authors were unable to locate research related to the evaluation of faculty teaching in field-based settings such as supervising teacher candidates in K-12 classroom clinical placements. As mentioned in the *Scholarship* section above, faculty in fields such as teaching

and nursing often have heavy teaching loads with significant time committed to supervising candidates in the field. Surveys used for student evaluation of faculty teaching are not designed to assess the quality of faculty teaching in field-based settings (Sidwell et al., 2025). This means that a significant portion of the teaching responsibilities for some faculty are not effectively evaluated.

Service

Service is the third domain in which faculty are commonly evaluated. Prior research on faculty evaluation in this area indicated a lack of clear definitions of what constitutes service and how performance in this domain can best be measured (Colbrek, 2002; Ward, 2003). Baker et al. (2015) found that universities often evaluate service in terms of the audience: service to department, service to college, service to university, and service to the broader community. However, they noted there is a lack of structured frameworks for evaluating faculty performance in each of these areas.

More recent research has examined inequities in how the service productivity of women and faculty of color are evaluated. Gerwin (2020) found that minoritized faculty are often overburdened by formal service commitments. They are frequently called upon to represent a diverse perspective at all levels. Informal service commitments also take their toll. Minoritized tenured and tenure-track faculty are sought out by minoritized students to assist them with navigating academia which these faculty often see as their duty. If these faculty are bilingual, they are often asked to translate in both formal and informal settings. Gerwin (2020) describes these additional burdens as a “time tax” that limits their ability to perform to the level of White colleagues who do not have this same burden.

Method

This case study employed a qualitative, process-oriented approach to document and analyze the revision of a faculty evaluation system within a merged academic department at a mid-size regional university. The aim was to describe and reflect on the development of a new, more equitable merit evaluation framework that aligned with evolving departmental priorities, structural shifts, and institutional goals. Given the complex organizational, interpersonal, and cultural factors involved, this study adopted a qualitative case study design (Yin, 2018). A single, bounded case - the development and implementation of a revised faculty evaluation system - was examined in depth to generate insights about institutional change processes, faculty engagement, and systems design.

Data were collected through participant observation, document analysis, and facilitated discussions, enabling triangulation of sources and perspectives (Yin, 2018). Primary data sources included meeting notes from the evaluation system design process, notes and written reflections from faculty meetings in which the evaluation system was discussed and critiqued, and faculty feedback from various points of the evaluation system process. Additional data was collected from the department chair's perspective of the previous evaluation system via document analysis, written reflections, and the meetings with the department chair and the faculty committee leading the redesign and development of the new evaluation system. These points included training for faculty on how to complete their evaluation materials, scorer training for faculty serving as peer-evaluators, and feedback from both the chair and faculty on the overall effectiveness of the evaluation system. Field notes recorded by the authors throughout the 18-month period described in this manuscript provided additional valuable information.

Case Context

In the five years leading up to the faculty evaluation system (FES) revisions described in this paper, the department underwent a series of significant structural and leadership changes that affected how faculty are evaluated and awarded merit. From an organizational standpoint, two smaller departments merged to form the current departmental structure. To do this, a department of about 25 tenured/tenure-track faculty focused on language, literacy, and special populations merged with a smaller department of about 15 tenured/tenure-track faculty focused on curriculum and instruction. Each department had previously had its own FES framework that was used for merit-based pay decisions. As a result of the merger, the two departments were required to merge their FES processes into a unified system that would now be used for approximately 50 tenured/tenure-track faculty. The most recent FES committee chairs from the two merging departments reached what was intended to be a short-term compromise on a joint FES framework that retained elements and criteria from each previous system. This joint framework was intended to be used for one year, during which time faculty and the department chair of the newly merged larger department would evaluate the effectiveness of the system and adjust accordingly.

However, that short-term FES evaluation system compromise extended for the next five years due to significant changes to departmental leadership. Over the next five years, the new department had three different department chairs. The first chair was in the position for one year. This department chair largely ignored the self-evaluation and peer-review components of the temporary evaluation framework, preferring instead to use her own opinion of each faculty member's performance. This was followed by an interim chair who served for two years while simultaneously serving in other leadership positions within the college. This interim chair took the opposite approach from their predecessor, deferring to each faculty member's self-evaluation

and paying limited attention to the peer-review elements of the temporary system. At this point, a full-time chair was hired from a different university system within the state that used different annual evaluation practices. After meeting with faculty across all ranks and program areas about the current evaluation system, the new chair was motivated to work with faculty to revise the FES framework so that it could more accurately evaluate the contributions of tenured/tenure-track faculty members in this new larger department.

Departmental priorities also underwent a significant change during this five-year period. At the undergraduate level, the teacher preparation program adopted an intensive, year-long residency model for student teaching. As part of this program, Undergraduate faculty provided intensive, field-based support to teacher candidates across their final four semesters in the program. Additionally, the department significantly expanded its capacity to offer programs that meet the needs of a diverse undergraduate student population, including first generation college students, students from minoritized and marginalized groups, and non-traditional students who are working and supporting families while simultaneously pursuing an undergraduate degree. The existing FES framework used during this time did not reflect the work now being prioritized by faculty. Instead, the existing FES framework focused on more traditional faculty responsibilities related to classroom teaching, research productivity, and service to professional organizations that is more readily accomplished by faculty teaching in graduate programs.

The new department chair identified this as an important issue to address during the first set of annual evaluations that she navigated. She noticed that the existing evaluation system was geared toward acknowledging teaching, scholarship, and service more closely aligned with graduate teaching responsibilities. This did not align with the department's unique make up of large undergraduate programs in which faculty engaged in developing and supporting new

teachers who were embedded in both urban and rural schools across the region. The previous FES framework valued traditional classroom teaching at the graduate level but lacked the acknowledgement and reward for the intense time commitment and investment required to support a diverse and changing undergraduate student population participating in intensive, field-based learning experiences. The new department chair hoped to work with faculty to develop a more equitable, representative structure to acknowledge the effort necessary to support and develop students at both the undergraduate and graduate levels.

The move to more field- and time-intensive undergraduate teacher preparation work by faculty coincided with the university's recent move to R2 status as a university with high research productivity. This was a final significant factor in the years leading up to the FES revision process described in this manuscript. The university's recognition of R2 status came from several years of increasing emphasis on pursuing external funding at the university, college, and departmental levels. Seeking and receiving external grant funding was a priority area consistently promoted by the university president and provost. Although conversations regarding how this new emphasis affected faculty evaluation had been ongoing, there had been no clear effort to accurately capture this work and its implications for faculty performance. External funding was only minimally considered in the existing FES framework, and there was still a lack of agreement among faculty on exactly what extent the pursuit of external funding should factor into faculty evaluation. The new department chair and faculty engaged in pursuing external funding identified this as an area that needed to be addressed in any revisions to the FES framework.

Process

An associate professor and a full professor volunteered to co-coordinate the performance evaluation process revision. Both were involved in the year-long residency program and had been heavily involved in the curriculum revision that had previously occurred at the undergraduate level. One taught exclusively in the undergraduate program, while the other taught at both the undergraduate and master's level. Ten committee members were purposefully selected based on representation of all program areas: curriculum and instruction, literacy, special education and bilingual/English as a second language. Equal representation at the full, associate and assistant levels, as well undergraduate, master's and doctoral faculty were additional criteria. The former process of evaluation often privileged faculty who taught doctoral courses, which tended to have fewer students and increased publishing opportunities with graduate students. Therefore, having equitable representation for all faculty was critical.

This resulted in a committee of 13: two co-chairs, the department chair, three full, three associate professors, and four assistant professors. Faculty representation was equitably distributed: one co-chair was a full professor, plus the three full professors on the committee; the second co-chair was an associate professor, plus the three associate professors on the committee, and so each rank had four members represented in the process. The committee was given the charge to design the evaluation system in accordance with university and college policies. The university policy was written in such a way as to allow for differences between college policies. Similarly, the college policy was written to allow differences between departments. The generic way the university and college policies were written had contributed to the differences in the two former departments whose merger was now leading to the FES system described in this paper. The generic language used in these policies also gave the committee the flexibility to design a system based on the specifics of the department. Once the committee developed the new

evaluation system, the plan had to be approved by all faculty and both college and university administrators.

Committee meetings were held during the summer and fall, initially biweekly. This committee developed a scoring structure that presented the totality of the teaching, scholarship, and service with effort to avoid privileging any specific program area based on discipline (e.g., privileging faculty in one program area over another) or by the degree levels offered to students (e.g., privileging faculty who might teach in undergraduate, masters, or doctoral level programs). Furthermore, it was critical to provide opportunities for assistant professors who often scored lower in the former system, particularly in service, because they did not have access to college or university committees and these committees had higher scoring values than the departmental committees that were more readily available to assistant professors. As a result, some faculty members engaged in minimal service to the department because it was not valued in the evaluation system. The new department chair was concerned about this because departmental committees and presence at events were critical to the department's functioning.

The entire committee discussed each category: teaching, scholarship, and service. Initially, it was difficult for committee members to see beyond what would be personally beneficial to them, particularly those who were already tenured. Advocating for one's own agenda wasted a lot of time. As a result, the co-chairs had to step in and narrowly define the subsequent meeting's parameters and ensure that assistant professors could participate in the discussion. Thus, most members could discuss what constituted each category from a holistic perspective that would be more inclusive of labor intensive, albeit less glamorous, responsibilities in the department and the college.

The committee solicited input from all faculty at departmental meetings with each committee member serving as a facilitator of small groups. This was critical with respect to providing agency and voice to the entire faculty. Certainly, personal agendas came to the forefront, but small group discussions enabled the committee to avoid whole group bedlam where the loudest voices would dominate. The co-chairs synthesized the data from the committee and small group discussions and brought back a framework that subdivided teaching, service, and scholarship with no one area of dominance in each category except for one required category in scholarship, publications, and one in service, departmental service. Rather than narrowly defining each level with a specific item like serving on the faculty senate, the co-chairs defined the categories of the rubrics in a general fashion, such as a once in a lifetime achievement or a labor-intensive teaching practice with high impact on student learning.

The co-chairs took the resulting rubrics in scholarship, service, and teaching back to the faculty for a vote. The rubrics were approved unanimously. Following the approval of the system, the co-chairs facilitated opportunities for faculty to provide examples of successful accomplishments at each level in the rubrics. The co-chairs took these examples provided by faculty members and created a mock faculty member scoring sheet that would be used to train scorers before the official evaluation in the spring. Throughout the final steps of this process, a header was included on all documents that contained the evaluative criteria and scoring examples that presented the results of the unanimous approval vote. This served as a reminder for faculty that they had reviewed and approved these new criteria. This was deemed necessary so that some faculty who were inclined to immediately begin critiquing the new system would be reminded that they had in fact voted in favor.

Lessons Learned

Although the new system addressed some of the issues noted in the literature, it could not overcome all of them. In scholarship, weight was given to a faculty members' particular research area by reducing the influence of impact factor and other metrics by which the quality of research is measured (Schimanski & Alperin, 2018; Settles et al., 2021). In teaching, the new system is an improvement upon the previous structure, as it recognizes and values the contributions of working in both field-intensive undergraduate licensure programs and graduate programs (Smith & Walker, 2022; Sidwell et al., 2025). In addition, weight was given to significant efforts to engage in professional development related to teaching. It did not, however, remove the bias inherent in students' evaluations of teaching (Aruguete, et al., 2017). In service, it placed departmental, college, university and professional service on equal footing. This assisted in alleviating some, albeit not all, of the time tax placed on minoritized faculty (Gerwin, 2020).

It is important to note that many faculty members were dissatisfied with the former process and therefore open to change. There were grumblings for several years from junior faculty and primarily undergraduate faculty about the unfairness of the evaluation system that privileged some and not others. In addition, the chair was receptive to change, particularly as it related to service. Work done at the program and departmental levels was not valued as much as that of the university, college, and profession, but it had to be done. Dissatisfaction is a great motivator of change.

Another of the most important lessons was to provide multiple opportunities for faculty input at various points in the process. Faculty in academia often do not hesitate to voice their opinions, some more than others. They were provided with the space to share their thoughts. Talking it out in small groups encouraged some faculty to see beyond their own work. To be

successful at developing an evaluation system, it is critical that faculty consider diverse accomplishments in service, scholarship and teaching rather than just their usual achievements. Certainly, not all faculty could make this shift, but most saw value in the work of others.

Notably, strong co-chairs are critical to the success of the process. The original committee of thirteen was unwieldly. Initial meetings devolved into chaos and those with the loudest voices dominated. The co-chairs eventually took control in subsequent meetings to narrowly define the tasks and ensure that the loudest voices were not the only ones to speak. They also chose to limit input from all faculty members to small group discussions.

Despite having a committee, the co-chairs had to do considerable work behind the scenes to synthesize all. Trying to engage the entire committee in developing the rubrics would have been impossible. There must be leaders of the process to develop the direction of the work. Fortunately, the co-chairs had a history of working toward the common good with respect to service, teaching, and research.

Conclusion

There are multiple opportunities for future research related to the design and implementation of this faculty evaluation system. At the university that served as the site for this case study, the authors intend to focus future research on the effectiveness and implications of the evaluation system for tenure-track faculty as compared to fully tenured faculty. For the system to be effective, it needs to fairly evaluate the work of faculty regardless of tenure status. Additionally, this university has a different evaluation system for assessing the performance of clinical (non-tenure track) faculty. There is ongoing discussion regarding the feasibility of replicating the evaluation system described in this manuscript for clinical faculty. This decision-making process, as well as the potential implementation of the system for clinical faculty, is an

area for future research. Finally, the authors are hopeful that faculty at another university will consider engaging in a similar revision process to their faculty evaluation system. Implementation and analysis at other university settings would provide valuable information regarding the design and implementation of faculty evaluation systems.

From a policy perspective, the work described in this case study highlights the importance of faculty taking a leading role in the design and implementation of their evaluation systems. This process was initiated by an associate professor and a full professor who volunteered to co-coordinate the performance evaluation process revision. They did so with the support of a department chair who recognized the need for improvements to the evaluation system. The chair then recommended an initial committee of thirteen faculty from a cross-section of program areas and ranks. This level of faculty voice and influence in the evaluation system was critical for ensuring buy-in and investment in the evaluation process. A key element to the revision and any future success that can be attributed to the policy revised in this case study is engaging faculty with administration, in this instance, the department chair, to examine existing policies. While the generic language evident in college and university policies allowed for flexibility, the university policy did place some constraining criteria on the evaluation system that could be developed. Without examining the college and university policies as part of the work, faculty understanding and buy-in is limited. Although the long-term success of this evaluation model is to be determined, empowering faculty to take a leading role in designing and implementing a robust evaluation system that is intended to address potential inequities is an important policy point for university leaders to consider.

This case study examined how a large department at a mid-size regional university reimagined its merit and faculty evaluation system to more equitably recognize undergraduate

contributions and better balance the evaluation of scholarship, service, and teaching. This case highlighted both the challenges and opportunities involved in creating merit-based evaluation systems that not only incentivize faculty productivity but also offer a fair and meaningful assessment of faculty work. As authors, we hope that detailing this process will provide readers with insights that can potentially inform the redesign of their own faculty evaluation systems.

References

Akins, L. & Murphy, L. (2019) Panel: Peer-based faculty evaluation v. student evaluation of teaching, *Journal of Collective Bargaining in the Academy, 0*(57). https://thekeep.eiu.edu/jcba/vol0/iss14/57.

Aruguete, M. S., Slater, J., & Mwaikinda, S.R. (2017). The effects of professors' race and clothing style on student evaluations. *The Journal of Negro Education*, *86*(4), 494–502. http://www.jstor.org/stable/10.7709/jnegroeducation.86.4.0494

Baker, D.F., Neely, W. P., Prenshaw, P.J., & Taylor, P.A. (2015). Developing a multi-dimensional evaluation framework for faculty teaching and service performance. *Journal of Academic Administration in Higher Education, 11*(2), 29-41.

Bichsel, J., Fuesting, M., Nadel-Hawthorne, S., & Schmidt, A. (2021). *Faculty in higher education annual report: Key findings, trends, and comprehensive tables for tenure-track, non-tenure teaching, and non-tenure research faculty and summary data for adjunct faculty for the 2020-21 academic year*. College and University Professional Association for Human Resources. https://www.cupahr.org/surveys/results/.

Buller, J.L. (2009, June). The pros and cons of merit pay. *Academic Leader*, *25*(6), 7-8.

Colbrek, C.L. (2002). Integration: Evaluating faculty work as a whole. *New Directions for Institutional Research*, *2002*(114), 43-52.

Drue, C.R., Bifulco, C.A. (2025). Beyond the numbers: How directors and chairs interpret student feedback to equitably evaluate teaching. *Journal of Academic Ethics*. Online first. https://doi.org/10.1007/s10805-025-09611-5

Field, L. (2015) Using outperformance pay to motivate academics: Insiders' accounts of promises and problems. *Australian Universities' Review*, *57*(2) 5-16.

Gannon, K. (2021). Faculty evaluation after the pandemic. *Chronicle of Higher Education, 67*(21) 1-6.

Gerwin V (2020, June 24). The time tax put on scientists of colour. *Nat Career Features.* https://doi.org/10.1038/d41586-020-01920-6

Hunnicutt, G.G., Lesher Taylor, L.R. & Keeffe, M.J. (1991). An exploratory examination of faculty evaluation and merit compensation systems in Texas colleges and universities. *Journal of the College and University Personnel Association, 42*(1) 13-21.

Laube, H., Massoni, K., Sprague, J., & Ferber, A. L. (2007). The impact of gender on the evaluation of teaching: What we know and what we can do. *NWSA Journal*, *19*(3), 87–104. http://www.jstor.org/stable/40071230

Miller, R.I. (1988). Merit pay in United States postsecondary institutions. *Higher Education, 17*(2) 219-232. https://doi.org/10.1017/S104909651800001X

Mitchell, K. M. W., & Martin, J. (2018). Gender bias in student evaluations. *PS: Political Science& Politics*, *51*(3), 648–652. https://doi.org/10.1017/S104909651800001X

Schimanski, L. A., & Alperin, J. P. (2018). The evaluation of scholarship in academic promotion and tenure processes: Past, present, and future. *F1000Research*, *7*, 1605. https://doi.org/10.12688/f1000research.16493.1

Settles, I. H., Jones, M. K., Buchanan, N. T., & Dotson, K. (2021). Epistemic exclusion: Scholar(ly) devaluation that marginalizes faculty of color. *Journal of Diversity in Higher Education, 14*(4), 493–507. https://doi.org/10.1037/dhe0000174

Sidwell, D., Lee, D., Zimmerman, P., Bentley, S., Barton, M. (2025). Teaching faculty experiences with student evaluation of instruction: A mixed-methods study. *Teaching and Learning in Nursing*, *20*(1), e276-e284. https://doi.org/10.1016/j.teln.2024.11.009

Smith, S., & Walker, D. (2022). Scholarship and teaching-focused roles: An exploratory study of academics' experiences and perceptions of support. *Innovations in Education and Teaching International*, *61*(1), 193–204. https://doi.org/10.1080/14703297.2022.2132981

Ward, K. (2003). Faculty service roles and the scholarship of engagement. *ERIC Digest*, ERIC Clearinghouse on Higher Education. https://files.eric.ed.gov/fulltext/ED480469.pdf

Williams, D. A. (2007). Examining the relation between race and student evaluations of faculty members: A literature review. *Profession*, 168–173. http://www.jstor.org/stable/25595863.

Yin, R.K. (2018). *Case study research and applications: Designs and methods* (6th ed.). Sage.