

# From mandate to co-create: leading the development of inclusive performance evaluation criteria

Jessi L. Smith

*Department of Psychology and Office of Research,  
University of Colorado Colorado Springs, Colorado Springs, Colorado, USA*

Sylvia Mendez

*Department of Leadership, Research, and Foundations,  
University of Colorado Colorado Springs, Colorado Springs, Colorado, USA*

Jennifer Poe

*Office of Research, University of Colorado Colorado Springs,  
Colorado Springs, Colorado, USA*

Camille Johnson

*Letters Arts and Social Science Dean's Office, Cal Poly Pomona, Pomona,  
California, USA*

Dale K. Willson

*Department of Leadership, Research, and Foundations,  
University of Colorado Colorado Springs, Colorado Springs, Colorado, USA*

Elizabeth A. Daniels

*University of the West of England, Bristol, UK*

Heather Song

*Department of Electrical and Computer Engineering,  
University of Colorado Colorado Springs, Colorado Springs, Colorado, USA, and*

Emily Skop

*Department of Geography and Environmental Studies,  
University of Colorado Colorado Springs, Colorado Springs, Colorado, USA*

## Abstract

**Purpose** – Annual performance evaluations of faculty are a routine, yet essential, task in higher education. Creating (or revising) performance criteria presents an opportunity for leaders to work with their teams to co-create evaluation metrics that broaden participation and minimise inequity. The purpose of this study was to support organisational leaders in developing equitable performance criteria.

© Jessi L. Smith, Sylvia Mendez, Jennifer Poe, Camille Johnson, Dale K. Willson, Elizabeth A. Daniels, Heather Song and Emily Skop. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

The project reported here was funded by the National Science Foundation (No. 2117351). Any opinions, findings and conclusions or recommendations expressed in this manuscript are those of the authors and do not necessarily reflect the views of the National Science Foundation. We are grateful to the West Virginia ADVANCE team for their collaboration and coaching and to Dr Kelly McNear for her feedback on a previous version.



---

**Design/methodology/approach** – We adopted the “dual-agenda” dialogues training that draws on concepts of collective self-efficacy and intersectionality for department leaders to co-create annual review criteria with their faculty members at one university. We used qualitative and quantitative data to assess the training and conducted an equity audit of the resulting annual review criteria.

**Findings** – Survey results from faculty members and departmental leaders (n = 166) demonstrated general satisfaction with the process used to create new criteria, perceptions that their criteria were inclusive and optimism about future reviews. Those with greater familiarity with the dialogues process had more positive perceptions of the inclusivity of their department’s criteria and more positive expectations of future reviews. The examination of eight indicators of equity illustrated that the resultant criteria were transparent and holistic.

**Originality/value** – This study builds on the relatively little research on faculty members’ annual performance evaluations, focussing on inclusive dialogues that centre equity and diversity. Results highlight the value of providing department leaders with evidence-based tools to foster system-level change through equitable evaluation policies. A toolkit is available for adaptation of the “dual-agenda” leadership training to both co-create annual review criteria and improve equity and inclusion.

**Keywords** Performance reviews, Leadership, Higher education, Collaboration, Inclusion, Faculty evaluation  
**Paper type** Research paper

Most professional workplaces include annual performance evaluations. At best, these evaluations are an opportunity for reflection, celebration, course correction and mapping future needs and goals. At worst, they are a space where hasty judgements, bias and distrust can significantly impact organisational culture, affecting employee productivity, job satisfaction, morale and career progression (Arvey and Murphy, 1998). Creating evaluation criteria can take many forms depending on the organisation’s structure. Within universities in the United States of America (USA), shared governance is the prevalent structure in which faculty members, students, staff and administrators jointly govern. Situated within this collaborative structure, we describe the implementation of a technique for department leaders to use with their faculty members to co-create annual performance evaluation policies. Through dialogue, leaders can align criteria with institutional goals, promote fairness and recognise diverse contributions.

Academic leaders often enjoy great agency to define the expectations for performance evaluations (Campbell and O’Meara, 2013; O’Meara *et al.*, 2015). Although unions or human resource departments may mandate some criteria, interpretation and implementation are frequently left to the department chair. In the USA, department chairs are often middle managers who must vacillate between sense-making of university expectations and sense-giving of those expectations to their faculty members (Hughes *et al.*, 2022). Moreover, though the principles of shared governance value faculty member collaboration in creating evaluation policies and processes, meaningful engagement of overtaxed faculty members in policy creation is difficult (Milligan, 2016). There is often little time for deliberate and inclusive policy work due to competing demands on faculty’s time (e.g. Misra *et al.*, 2012), which made worse during the global pandemic (e.g. Krukowski *et al.*, 2021). Therefore, one advantageous aspect of the current approach was the use of inspirational change strategies that emphasised the greater purpose and opportunities ahead whilst staving off a corrosive focus on the perceived obstacles associated with policy change (Bass and Riggio, 2006).

### **Equity, inclusion, transparency and performance evaluations**

Within USA higher education contexts, performance evaluations of faculty members can take several forms. For example, formal evaluations for faculty members on track for tenure, retention or promotion typically involve decision-making at multiple levels, including peer evaluation. Other evaluations, such as those for merit pay, may fall solely on the chair. No matter the type of evaluation, the criteria not only convey expectations for job performance

but also communicate the values of the institution (Hardré and Cox, 2009). Performance evaluation processes also have the potential to maintain the status quo of white, male, heteronormative definitions of merit (Blair-Loy and Cech, 2022). Historically, universities were established by and for the privileged class (Bonilla-Silva and Peoples, 2022). As a result, universities today are often gendered and racialised organisations rife with structural inequalities despite explicit commitments to diversity, equity and inclusion (Kaiser *et al.*, 2013; Niemann *et al.*, 2020). Such organisational structures, including performance evaluation criteria, are part of the complex explanation for why racial minority and women-identifying faculty members do not progress in their careers as often as their white, male counterparts. For example, in the USA, there are fewer racial minority and women-identified faculty than proportionally expected as full professors (e.g. Colby and Fowler, 2020), and they are not often in top academic leadership positions (e.g. Fox Tree and Vaid, 2022; Whitford, 2020).

Faculty member evaluation policies are key to creating an organisational culture aligned with a university's mission (O'Meara, 2005). For example, service work by faculty members is often not highly valued in evaluation, even though it is essential to the university's functioning. Moreover, service is an equity issue because women and faculty members of colour perform much of this time-consuming and emotionally intensive service work (e.g. Järvinen and Mik-Meyer, 2024). Although policy alone cannot change an entrenched academic culture (e.g. Barratt-Pugh and Krestelica, 2019), inclusive and equitable tenure and promotion processes are critical to systematic transformation (Mitchneck *et al.*, 2016). For these reasons, most research has centred on retention, promotion and tenure policies (e.g. Cate *et al.*, 2022; O'Meara, 2005). However, we suggest annual performance evaluations are an equally important lever for change. For example, where annual performance evaluations are used to determine merit pay, inequitable policies may damage the financial standing of minoritised and marginalised faculty members and contribute to pay gaps (e.g. Casad *et al.*, 2022).

We draw from well-developed research on promotion/tenure evaluation structures to inform our innovative approach to leading inclusive and equitable annual faculty-member performance evaluation criteria construction. We used the American Council on Education's (ACE) report on equity-minded promotion/tenure reform (O'Meara *et al.*, 2022), which outlines the importance of holistic assessment using the principles of "transparency, clarity, accountability, context, credit, consistency, flexibility, agency, and representation" within the context of academic work related to scholarship, teaching, service and institution building. With these foci in mind, we adopted the evidence-based dialogues dual-agenda technique to engage department chairs in the work of co-creating new annual performance evaluation criteria with their faculty members.

### **The dialogues dual-agenda technique**

The dialogues dual-agenda process was developed at West Virginia University (WVU) as part of their National Science Foundation (NSF)-funded ADVANCE program (Latimer *et al.*, 2014). The "dual-agenda" approach intends to deliver on two goals: (1) completing a department-level task (e.g. writing a strategic plan) and (2) enhancing inclusion, equity and perceptions of fairness (Latimer *et al.*, 2014). This "win-win" dual-agenda technique emphasises a "co-creation" approach and supports open conversations in which minoritised and marginalised voices are brought to the forefront, thereby simultaneously creating a positive work climate that supports equity and diversity and completing the assigned department's task (Bird and Latimer, 2019). The facilitation-based process is underpinned by the power of collective self-efficacy – a group's shared understanding that leveraging members' talents and capacities to co-create processes leads to optimal outcomes (Bandura, 1997). The advantage of this co-creation technique is that it strengthens departmental relationships

---

amongst diverse faculty members and department chairs by generating trust, establishing consensus on goals and catalysing action by demonstrating inclusive techniques, practicing good communication strategies and making intentional space for everyone to participate and reflect (Holmes *et al.*, 2016). The result of co-creation through dialogues is improved gender dynamics and increased equal opportunities (Bird and Latimer, 2019; Latimer *et al.*, 2014). Dialogues are especially effective in majority-male departments (Latimer *et al.*, 2014) and has been used at other USA universities focussed on a range of topics, including Institutional Review Board (IRB) processes, strategic planning, promotion/tenure criteria and hiring (Jackson *et al.*, 2023). To the best of our knowledge, we are the first to adapt this dual-agenda dialogues programme to generate annual performance evaluation criteria. Our adaptation included a unique focus on department leaders working in fields with historical underrepresentation of women-identified and faculty of colour in science, technology, engineering and mathematics (STEM) and social and behavioural sciences (SBS), with the intent to scale to other department leaders at one USA university.

### **Project overview**

Our goal was to examine the subjective experiences and outcomes of the adapted dialogues process for department leaders and faculty members to co-create annual performance evaluation criteria. The need for such criteria was driven by a new Board of Regents policy mandating the separation of annual reviews from promotion/tenure criteria. Until this point, no unit had any stand-alone criteria; department leaders thus faced the challenge of creating criteria from scratch. Taking advantage of the timing for the mandated policy work, the project unfolded in three phases: (1) train department leaders in the adapted WVU ADVANCE Dialogues programme and create a toolkit to adopt the process for their own unit's use; (2) survey faculty members and departmental leaders on their experience of the process and opinions on their created criteria and (3) code the generated criteria documents for transparency, accountability, context, holistic assessment and broad considerations of faculty work with a rubric inspired by the ACE (2022) audit resource.

### **Phase 1 methods: leadership training**

#### *Participants*

A total of 14 department leaders (of 15 possible) in STEM/SBS departments across two colleges engaged in the dialogues leadership training (described below) over four days, with 12 completing the training. All but two of the department leaders were the current chairs of the unit. Participants were primarily male (78.6%) and white-identifying (64.3%).

#### *Procedure*

Partnering with WVU, we adapted the dialogues training to a virtual format, given pandemic concerns. The high-engagement training began with nine hours of virtual engagement over three days before the start of the fall 2021 semester. The training used a comprehensive and interactive flexible facilitation technique that allowed for anonymous, iterative feedback and group-level discussions about annual performance criteria, review processes and how evaluation connects to an inclusive department culture. In this way, the dual agenda supported the construction of the inaugural annual review criteria and demonstrated co-creation strategies to improve equity, diversity and inclusion.

Each session began with ground rules and a connect activity (e.g. pick a virtual button and discuss why you chose it), followed by a warm-up brainstorm (e.g. how many uses can you think of for a pencil?). Next, leaders engaged in individual brainstorming around a given question centred around the topic of annual review criteria and process (e.g. "How have

research practices changed in your subfield over the last 25 years?”, “How is service and mentoring related to equity, diversity, and inclusion recognized?”). We ensured the interactive questions were attentive to intersectionality to make visible how criteria can harm or privilege people with different positionalities (e.g. Hill Collins and Bilge, 2016). We also followed the “humble inquiry” approach to wording questions (Schein and Schein, 2021). Participants were put into randomly assigned breakout groups to generate themes that were then reported to the larger group for discussion. Each three-hour session was organised around seven steps (Figure 1).

The internal facilitation team consisted of five tenured faculty members, with logistical support provided by the programme director of Project CREST – an NSF-funded local programme aimed at creating an inclusive culture of research. The internal facilitation team took turns leading each step (see Figure 1) and assisting in the breakout rooms. They helped participants identify existing equity barriers and biases within their current understanding of annual review processes whilst explicitly focussing on how the annual review should consider the downstream impacts of COVID-19 on research productivity, especially for women faculty and faculty of colour.

The final stage was for the department chairs to “practice” their new skills with other academic affairs leaders. All told, 37 campus leaders participated in the 90-min virtual practice session. Following the session, the trained department chairs debriefed as a group. They then received a link to an anonymous open-ended dialogues survey as well as a certificate and a small gift card as a token of appreciation. We then created a toolkit (<https://tinyurl.com/annualreviewtoolkit>) for chairs to reference, which is freely available for use and

Step	Goal
1: Orient	Establish ground rules and communicate agenda
2: Connect	Form connections between members participating in the discussion
3: Warm-up	Practice brainstorm idea generation with an interactive activity
4: Individual Brainstorm	Focus on the specific agenda topic (annual review criteria in this case) and allow everyone to individually and anonymously brainstorm responses to key questions. The idea is to allow a free exchange of ideas and prompt creativity without worrying about backlash
5: Collaborate	In small groups, organize ideas by themes that emerge in the brainstorm. Allowing smaller groups to discuss and organize allows more collective voices to be heard and unique ideas to emerge
6: Report out	Based upon the themes, determine as a group what elements are most relevant and important to include in the annual review criteria
7: Commit to Action	Have each member take ownership of at least one action item they will address to ensure the session produces tangible results. This could be writing up a section, collecting more information, or organizing the meeting notes

**Figure 1.**  
Dialogue steps for  
department leader  
training session

**Source(s):** Figure by authors

---

includes information on essential steps for engaging leaders in the “dual-agenda” training, sample agendas for different levels of engagement and links to other tools and technologies for leaders to facilitate the co-creation of important departmental work.

### *Results and discussion*

Qualitative analysis of the responses illustrated that the dialogues process was positively received. Exemplar responses to the item “in what ways have you felt heard during this dialogues process?” included “always felt appreciated and heard,” “very open and inclusive process allowing anonymous and visible contributions” and “all could participate.” Exemplar responses to the item “what did you take away from this dialogues training?” included “these are great ideas for brainstorming in an inclusive manner in the future,” “there is a need for systemic change in processes and support for annual review as a developmental process” and “deeper engagement would be helpful and let more perspectives be fully explored.” Exemplar responses to the final item “how can you use this network of colleagues to support you moving forward in creating annual review criteria?” included “I’ll bring insights and process this effort back to my department as we move towards annual reviews this spring,” “work with both other chairs and deans to make it more collaborative and transparent” and “I appreciate knowing that others on campus are struggling with the development of new criteria but [am] finding this opportunity to change reviews to an inclusive process has a lot of potential.”

Data show the response to dialogues was positive and resulted in a commitment to co-creating new annual performance evaluation criteria over the coming year. At the end of that year, when the criteria materials were finalised, faculty members and chair perceptions of the co-creation process and expectations for how the new criteria would be used in the future were assessed via an online survey (see Phase 2 below).

## **Phase 2 methods: perceptions of the creation process and the criteria**

### *Participants*

Tenure-track faculty members and department chairs ( $n = 290$ ) were invited to complete an online survey to provide “feedback on the annual review creation/revision process.” A total of 149 ( $n = 134$  faculty members and  $n = 15$  department chairs) took the survey (response rate = 51.4%). Participants completed a consent form prior to participation and received a \$25 e-gift card as a token of appreciation. The university’s institutional review board approved the survey.

### *Procedure*

Participants were given an anonymous link to either a faculty member or a department chair survey on Qualtrics. The items were similar across the surveys, but instructions emphasised “leading” the process vs “participating” in the process. Survey items were counterbalanced, and when possible, we modelled items after published work-organisation measures (Deci *et al.*, 2001). The complete survey code book is available upon request. Here, we describe four constructs of interest. First, we assessed *satisfaction with the process* using four items (Cronbach’s  $\alpha = 0.926$ ) on a 1 (strongly disagree) to 7 (strongly agree) Likert scale; sample item “In general, I am extremely happy with the process we used to create/revise our unit’s annual review criteria.” Second, we assessed *perceptions of the inclusiveness of the criteria* using five items ( $\alpha = 0.827$ ) on a 1 (strongly disagree) to 7 (strongly agree) scale; sample item “Our department considered how the review criteria reflect the contributions of faculty from multiple social groups and backgrounds.” Third, we measured *future expectations about future annual reviews* with six items ( $\alpha = 0.905$ ) on a 1 (strongly disagree) to 7 (strongly agree)



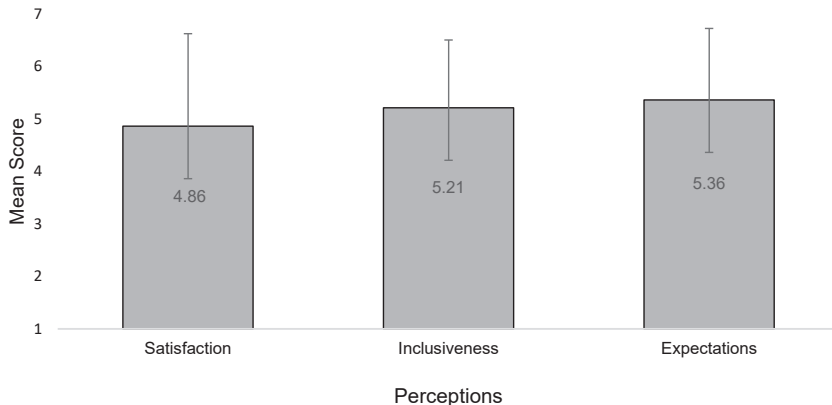
scale; sample item “I feel confident that next year’s annual review process will be fair.” Finally, we gauged *familiarity with the dialogues process* using a single item “How familiar are you with the dialogues technique from WVU?” using a 1 (not at all familiar) to 4 (very familiar) scale. We also asked general logistic questions, including “How often do faculty want to revisit the annual review criteria documents?” with a forced choice of once every 1, 2, 3, 4 or 5+ years. In addition, participants were asked to provide demographic data and information about the type of department they were in (e.g. STEM field).

*Results and discussion*

Our analytic strategy was as follows: (1) present descriptive patterns of the perceptions of the annual criteria process, (2) explore similarities or differences amongst department leader and faculty perceptions and (3) gauge associations between familiarity with the dialogues dual-agenda training/toolkit and inclusiveness of the criteria produced and future expectations of annual reviews.

First, our descriptive analyses illustrated in [Figure 2](#) indicate that, on average, participants were satisfied with the process of creating annual review criteria documents, perceived their documents to be inclusive and reported positive expectations for future annual reviews. To probe these patterns more systematically, we used one-sample *t*-tests to test the mean responses against the specific “neutral” scale point value. The average measure of *satisfaction with the process* used to create annual reviews ( $M = 4.86$ , standard deviation ( $SD$ ) = 1.76) was significantly higher than the scale midpoint ( $M = 4.00$ ),  $t(82) = 4.47$ ,  $p < 0.001$ . The average measure of the *perceived inclusiveness of the criteria* ( $M = 5.21$ ,  $SD = 1.29$ ) were also significantly higher than the scale midpoint,  $t(94) = 9.05$ ,  $p < 0.001$ . Finally, *positive expectations about future annual reviews* ( $M = 5.36$ ,  $SD = 1.36$ ) was significantly higher than the scale midpoint,  $t(95) = 9.74$ ,  $p < 0.001$ . Together, these findings indicate that participants’ perceptions of the annual review process were more positive than would be expected compared to neutral.

Next, we explored similarities and differences between the perceptions of the faculty members and the department chairs. Results showed that department chairs ( $M = 5.66$ ,  $SD = 0.92$ ) were significantly more satisfied with the document creation process compared



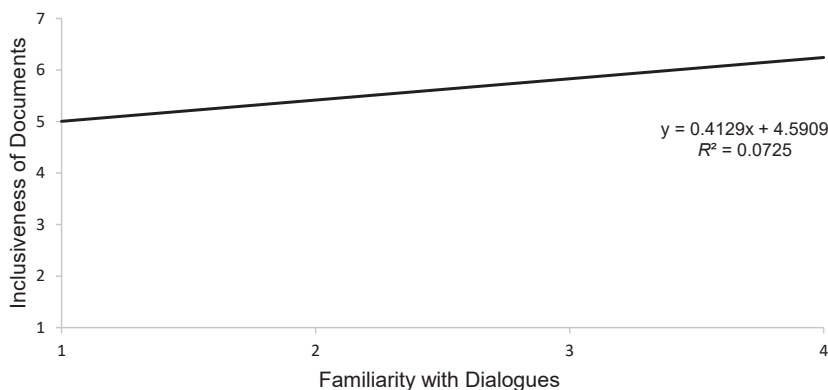
**Figure 2.** Perceptions of annual review criteria for document creation

**Note(s):** Higher scores indicate greater satisfaction with the process used to create the documents, perception of the documents as more inclusive, and more positive expectations about future annual reviews

**Source(s):** Figure by authors

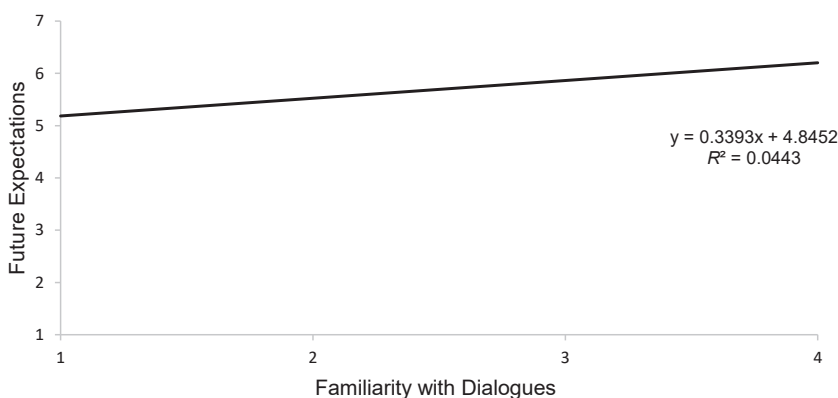
with faculty members ( $M = 4.74$ ,  $SD = 1.83$ ),  $t(81) = 9.11$ ,  $p = 0.015$ . Department chairs also perceived the documents to be more inclusive ( $M = 6.28$ ,  $SD = 0.64$ ) compared with faculty members ( $M = 5.04$ ,  $SD = 1.31$ ),  $t(93) = 9.25$ ,  $p < 0.001$ . Finally, department chairs ( $M = 6.45$ ,  $SD = 0.49$ ) reported more positive expectations about future reviews compared with faculty members ( $M = 5.18$ ,  $SD = 1.03$ ),  $t(94) = 13.23$ ,  $p < 0.001$ . Together, these findings indicate that department chairs had more positive attitudes about the annual review process than other faculty members. However, given that our sample was necessarily unequally distributed because there are far fewer chairs than faculty members, we take caution in interpreting these results.

Finally, we wanted to gauge associations between familiarity with the *dialogues technique*, the perceived inclusiveness of the created documents and expectations for future reviews. As shown in [Figures 3 and 4](#), the results of the regression analyses showed a significant positive relationship between familiarity with the dialogues training and positive perceptions of inclusiveness for the created criteria. Moreover, results also showed a significant positive relationship between familiarity with dialogues and positive expectations for future annual



**Figure 3.** Relationship between familiarity with the dialogues technique and the perceived inclusiveness of the annual review documents created

Source(s): Figure by authors



**Figure 4.** Relationship between familiarity with the dialogues technique and positive expectations for future reviews

Source(s): Figure by authors



reviews using the new criteria. Because we used a single item to measure familiarity, associations should be interpreted cautiously. Nevertheless, the results point to a promising impact of the leadership training on co-creating inclusive criteria. Of course, the actual criteria themselves are essential indicators of equity and transparency. Accordingly, we next audited review criteria based on eight indicators of equity.

## 96 Phase 3 methods: an audit on the quality of the created criteria

### Data collection

Annual review criteria documents in 29 departments were available for examination in 5 of the 6 academic colleges at the university. The remaining college was given an extension and was not included in this study.

### Procedure

To explore the way equity-minded practices were embedded in the inaugural annual review documents, a qualitative coding analysis was conducted using *The Project CREST Annual Review Coding Rubric for Equity* (Figure 5, see also <https://tinyurl.com/annualreviewtoolkit>), which was inspired by the *Equity-Minded Reform of Faculty Evaluation Policies Audit*

Annual Review Criteria Coding Rubric for Equity	
Scoring: 0 = Absent    1 = Addressed but Room for Improvement    2 = Fully Addressed	
<b>1: Transparency (M = 1.41)</b>	
1.1	The criteria are easily understood (i.e., no ambiguity could invite bias or misinterpretation).
1.2	The criteria are well-organized such that what meets expectations, exceeds expectations, and is outstanding is clear.
<b>2: Accountability (M = 0.89)</b>	
2.1	The criteria identify an appeal process.
<b>3: Context (M = 0.93)</b>	
3.1	The criteria provide ways to bring relevant life contexts into the review (e.g., significant transitions/ disruptions).
3.2	The criteria provide ways to bring relevant work contexts into the review (e.g., differentiated workloads, Sabbatical, leave of absence).
3.3	The criteria account for expected professional and collegial behavior.
3.4	The criteria consider faculty rank to determine meeting expectations, etc., in teaching/research/service.
3.5	The criteria include processes for faculty to explain the significance of their work.
<b>4: Holistic Assessment (M = 1.41)</b>	
4.1	The criteria are welcoming and open to multiple ways of conducting and engaging in teaching/research/ service.
4.2	The criteria promote a holistic assessment of teaching/research/service, rather than an overreliance on numeric metrics (e.g., FCOs, impact factor of journals, amount of grant dollars).
4.3	The criteria consider the ways in which faculty work contributes to the overall mission and goals of the unit/college/university across teaching/research/service.
4.4	The criteria recognize the different demands of a faculty member's type of teaching/research/service (e.g., interdisciplinary, collaborative, international, community-based).
4.5	The criteria value and recognize diversity, equity, and inclusion work within teaching/research/service.
4.6	The criteria acknowledge an achievement award is optional, but not necessary, to receive a rating of outstanding in teaching/research/service.
4.7	The criteria include an explicit statement about using the annual review process to map out future goals.
<b>5: Teaching Considerations (M = 1.06)</b>	
5.1	The criteria account for faculty engaged in different modes of instruction (e.g., in-person, online, hybrid).
5.2	The criteria recognize innovation in teaching (e.g., development of new courses, use of high-impact practices).
5.3	The criteria account for the availability of a teaching assistant, class size, class level, and the number of new prep in a given year.
5.4	The criteria indicate the evaluation of teaching includes a peer-review process, and the process is articulated.
5.5	The criteria recognize faculty engaged in student advising/mentoring informally and formally (e.g., lab supervision, independent study, thesis/dissertation membership).
5.6	The criteria seek student input to evaluate advising/mentoring quality.
5.7	The criteria consider writing student reference/recommendation letters as a valuable teaching activity.
<b>6: Research and Creative Works Considerations (M = 1.14)</b>	
6.1	The criteria provide a broad definition of research and creative works.
6.2	The criteria recognize a broad set of research and creative works products over specific types.
6.3	The criteria recognize faculty who publish on pedagogy-related topics as scholarship.
6.4	With the expansion of pay-to-publish outlets and predatory journals, the criteria clearly state peer-reviewed publications are expected.
6.5	The criteria recognize alternative products and venues for dissemination of scholarship (e.g., podcasts, blogs, social media, webinars).
6.6	The criteria recognize the value of varied publication outlets, venues, and presses to reach a broad audience.
6.7	The criteria define authorship expectations (e.g., author order, number of co-authors, publishing with an advisor).
6.8	The criteria explicitly value collaborative work with fellow faculty.
6.9	The criteria explicitly value collaborative work with undergraduate and/or graduate students.
<b>7: Grant Activity Considerations (M = 0.93)</b>	
7.1	The criteria recognize grant submissions along with grant awards.
7.2	The criteria recognize PI, Co-PI, evaluator, and consultant roles.
7.3	The criteria recognize the time and effort of administering, carrying out, and closing out grants.
7.4	The criteria specify the type of funding that is valued (e.g., external vs. internal, award amount, research vs. teaching vs. workforce development).
<b>8: Service and Leadership Considerations (M = 1.23)</b>	
8.1	The criteria articulate the value of campus service, define it, and delineate what is considered an appropriate amount.
8.2	The criteria articulate the value of disciplinary service, define it, and delineate what is considered an appropriate amount.
8.3	The criteria value high service levels to the unit/college/university.
8.4	The criteria value leadership over membership in service roles.
8.5	The criteria value efforts related to mentoring faculty within and outside the university.
8.6	The criteria account for compensated service work.

**Figure 5.**  
The annual review criteria coding rubric for equity, with overall construct mean scores

Source(s): Figure by authors

*Resource* published by the ACE in 2022. The *Audit Resource* outlines principle- and policy-related questions institutions should consider when creating equitable faculty evaluation policies and procedures. The rubric attends to transparency, accountability, context, holistic assessment and considerations of teaching, research and creative works, grant activity and service and leadership in the annual review guidelines and criteria.

Three coders – a professor external to the university, an advanced graduate research assistant and one of the internal facilitators – independently examined and coded all 29 annual review documents using the adapted audit tool in [Figure 5](#). The *a priori* goal was to have people with different perspectives read and interpret the documents to strengthen our methods. Amongst coders, the external faculty member possessed over a decade of experience evaluating faculty and was asked to think like an external evaluator. The faculty member internal to the university had evaluated faculty as a previous department chair and was asked to think like a chair. The advanced graduate student coder was asked to think like a brand-new tenure-track faculty member. Each coder was directed to approach this project from their unique point of view (versus the goal of being highly correlated) to achieve higher levels of external validity.

To familiarise the coders with the process, they first examined two annual review documents as practice and then compared their scores and discussed differences in coding to identify any differences in definitions of terms. As shown in [Figure 5](#), the rubric included 41 statements concerning transparency, accountability, context, holistic assessment and considerations of teaching, research and creative works, grant activity and service and leadership. Statements were scored from 0 = absent, 1 = addressed but room for improvement and 2 = fully addressed, with higher scores indicating greater equity-minded practice.

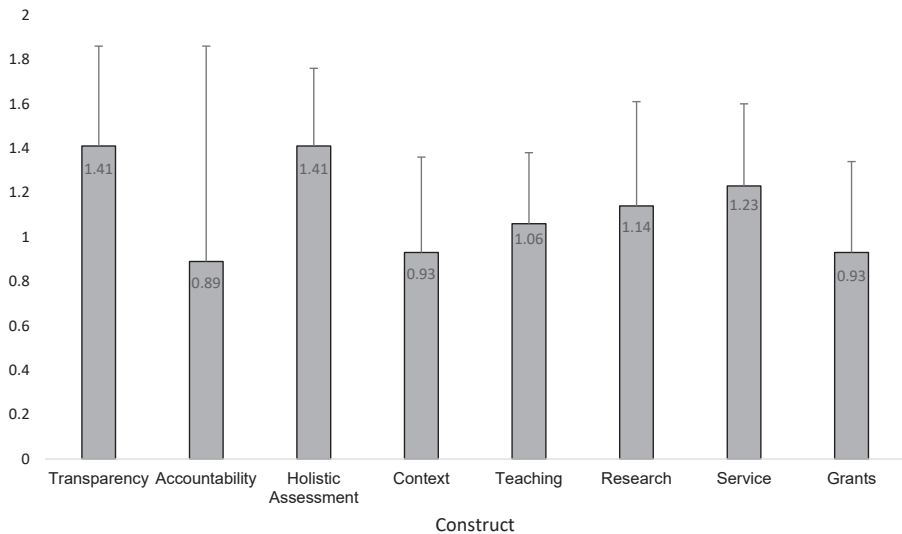
### *Results and discussion*

*Coding reliability.* Though our *a priori* goal was to have different vantage points review the criteria (versus trying to get agreement across all coders), we still calculated the inter-rater reliability using a two-way mixed consistency, average-measures intraclass correlation (ICC). The resulting overall ICC was in the very good to excellent range,  $ICC = 0.893$  ([Cicchetti, 1994](#)), indicating coders generally had a high degree of agreement and suggesting the annual review documents were, by and large, rated similarly across coders. Interestingly, the coders had very different perspectives on the construct of transparency ( $ICC = 0.15$ ), whereas all other construct ICCs were in the range of 0.61–0.97, which is considered good to excellent in agreement. We believe this finding is valuable when considering how different career stages can lead to varying understandings of transparency. What one person might see as obvious, another might see as oblique.

*Descriptive results.* The mean and SD for each construct for the university are reported in [Figure 6](#). The results of the qualitative coding analysis are discussed in descending order of mean ratings and emphasise the items under each construct that were rated the highest. Example criteria from our coding for each construct are presented in [Figure 7](#) and a link to all units' criteria is available in our toolkit.

Overall, the strongest equity-minded practice occurred in the areas of transparency ( $M = 1.41$ ,  $SD = 0.45$ ;  $t(87) = 8.61$ ,  $p < 0.001$ ) and holistic assessment ( $M = 1.41$ ,  $SD = 0.35$ ;  $t(87) = 10.78$ ,  $p < 0.001$ ), which were both significantly higher than the scale midpoint of 1. Regarding transparency, the coders viewed the annual review documents as generally understandable and well organised for faculty members to determine the distinction between “meeting expectations”, “exceeding expectations” and “outstanding”.

Under holistic assessment, the highest-rated criteria indicated that departments were open to counting a variety of ways in which faculty members conduct their work. Holistic



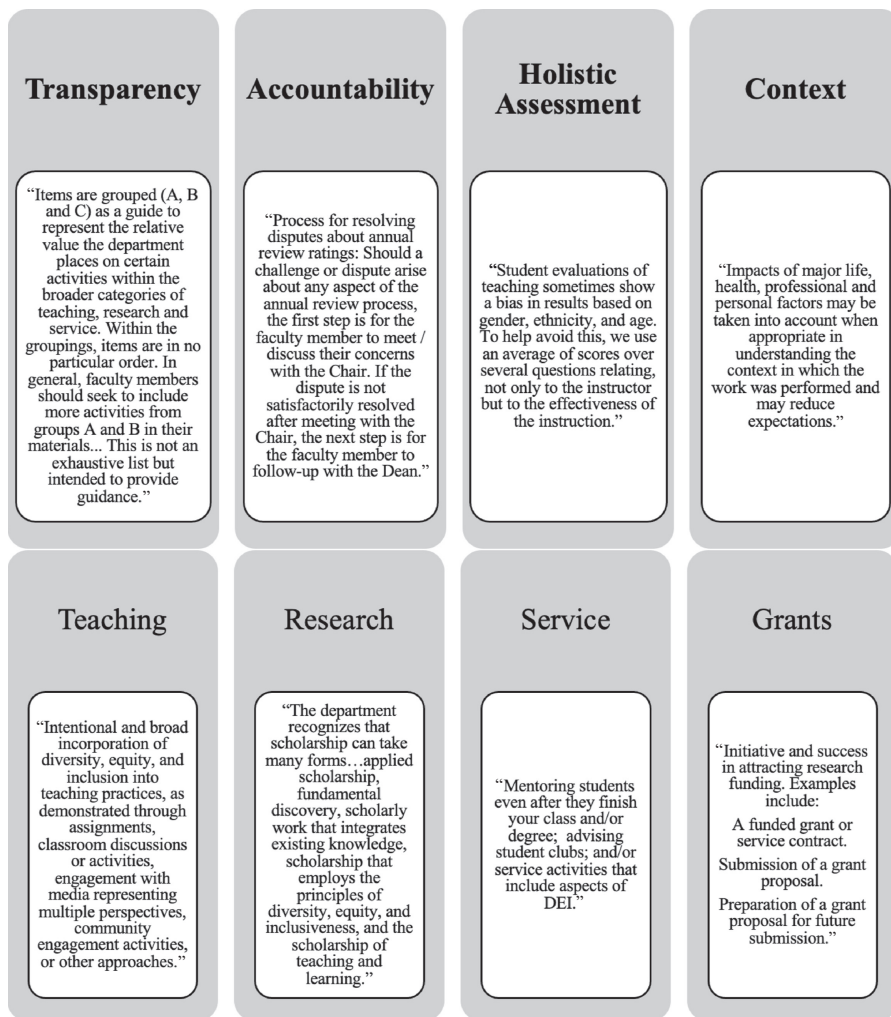
**Figure 6.**  
Annual review coding  
rubric results by  
construct for campus

**Note(s):** The average rating from the three coders is presented  
**Source(s):** Figure by authors

criteria also did not appear to overvalue numeric metrics (e.g. journal impact factors and student evaluation ratings) or achievement awards. All told, the results suggest transparency and holistic assessment were the two constructs most present in the newly created criteria.

Results of the coding showed that service and leadership considerations ( $M = 1.23$ ,  $SD = 0.37$ ), research and creative works considerations ( $M = 1.14$ ,  $SD = 0.35$ ) and teaching considerations ( $M = 1.06$ ,  $SD = 0.32$ ) were all rated as addressed, but there was room for improvement. For service and leadership, most departments strongly valued service to the campus and their disciplines and recognised the importance of leadership over purely membership on committees. Regarding research and creative works considerations, criteria recognised a broad set of research and creative works products and expected peer-reviewed publications. Finally, for teaching considerations, department criteria valued innovation in teaching and tended to consider peer-review ratings as an essential component to evaluate teaching.

Grant activity considerations ( $M = 0.93$ ,  $SD = 0.41$ ) were coded on the lower end of equity with much room for improvement. The highest-rated item under this construct indicated departments tended to recognise grant submissions along with awards but with little information on the valuation of different stages (proposal vs post-award) and roles (PI vs co-PI vs evaluator) of sponsored programme engagement. Context was coded similarly low ( $M = 0.93$ ,  $SD = 0.43$ ), with the highest-rated item indicating that faculty were given the opportunity to explain the significance of their work during the annual review process but with little attention paid to other important types of contextual issues such as how evaluations are considered when faculty experience a significant life event or are on sabbatical leave. Accountability ( $M = 0.89$ ,  $SD = 0.97$ ) was rated relatively lowest amongst all eight areas, partly due to most departments failing to identify appeals processes within their documents. However, this construct was not significantly different from the midpoint of the scale, suggesting it was just relatively lower than the other constructs.



**Source(s):** Figure by authors. All criteria are available through <https://research.uccs.edu/creating-annual-reviews> or via our toolkit at <https://tinyurl.com/annualreviewtoolkit>

**Figure 7.**  
Exemplar annual  
review criteria by  
construct

In summary, the audit of the annual review criteria documents showed that, by and large, the documents conveyed transparency and considered equity, diversity, and inclusion (EDI) and holistic assessment, with some room for improvement in specificity within specific areas of faculty members' work (e.g. service and grant activity), for process information related to accountability and taking into context a given faculty member's workload.

### General discussion

Annual performance evaluations are a hallmark of most professional workplaces. In some organisations, criteria are determined by top-down processes, whilst in others, criteria are

developed in collaboration with peers, managers and direct reports. In three phases, the current project described the implementation of the dialogues training programme designed to help department chairs work with fellow faculty members to develop equitable, inclusive and transparent annual review criteria and then assessed perceptions of the resulting training and annual review criteria. First, we examined the efficacy of implementing the dialogues facilitation process with department leaders. Using a train-the-trainers methodology, most department chairs in STEM and SBS at one university completed 9 h of training and then showcased what they learnt with other chairs and campus leaders. A toolkit (<https://tinyurl.com/annualreviewtoolkit>) was developed with the leadership training procedures and shared broadly with the campus to facilitate using dialogues-inspired processes.

Results revealed that department leaders viewed the training experience positively. Moreover, survey results demonstrated that faculty members also positively experienced the co-creation and held positive expectations for using the criteria in future annual reviews. Across the board, more familiarity with dialogues (e.g. through toolkit usage, involvement in training sessions or the co-creation process itself) was positively associated with greater perceptions of inclusivity of the newly created documents and with positive expectations for future reviews. An audit of the created documents showed that created documents were high in transparency and holistic assessment.

The goal of this project was to forge a new inclusive vision of policy co-creation that is mindful of and responsive to the needs and desires of both today's diverse faculty members and the next generation. With the bulk of performance appraisal research occurring in industry settings (e.g. Kuvaas, 2006), the current project builds on the relatively little research on leading change within faculty members annual performance evaluations (e.g. Barclay and York, 2003). By drawing on concepts of collective self-efficacy and intersectionality, the adapted WVU dialogues facilitation is a process through which department leaders can use our open-source toolkit (<https://tinyurl.com/annualreviewtoolkit>) for adopting inclusive, evidence-based approaches in service of organisational transformation that broadens participation and minimises bias and inequity in annual review policymaking.

As leaders and change agents collectively work towards equity in performance evaluations, we cannot lose sight of the individuals behind the process. We must remember that a faculty member's record often reflects the opportunities available to them and not their work ethic or quality of work (e.g. see Barclay and York, 2003). Using an asset-based lens for the evaluation of faculty work is imperative (e.g. O'Meara, 2005; O'Meara *et al.*, 2022). We believe that providing leaders the guidance and support they need to foster system-level change through co-created equitable evaluation policies is critical to the health of any organisation and is especially vital for institutions of higher education.

## References

- Arvey, R.D. and Murphy, K.R. (1998), "Performance evaluation in work settings", *Annual Review of Psychology*, Vol. 49 No. 1, pp. 141-168, doi: [10.1146/annurev.psych.49.1.141](https://doi.org/10.1146/annurev.psych.49.1.141).
- Bandura, A. (1997), *Self-Efficacy: the Exercise of Control*, W.H. Freeman and Company, New York, NY.
- Barclay, L.A. and York, K.M. (2003), "Clear logic and fuzzy guidance: a policy capturing study of merit raise decisions", *Public Personnel Management*, Vol. 32 No. 2, pp. 287-299, doi: [10.1177/009102600303200208](https://doi.org/10.1177/009102600303200208).
- Barratt-Pugh, L.G. and Krestelica, D. (2019), "Bullying in higher education: culture change requires more than policy", *Perspectives: Policy and Practice in Higher Education*, Vol. 23 Nos 2-3, pp. 109-114, doi: [10.1080/13603108.2018.1502211](https://doi.org/10.1080/13603108.2018.1502211).
- Bass, B.M. and Riggio, R.E. (2006), *Transformational Leadership*, Psychology Press, New York, NY.

- Bird, S. and Latimer, M. (2019), "Examining models of departmental engagement for greater equity: a case study of two applications of the dual agenda approach", *Equality, Diversity and Inclusion: An International Journal*, Vol. 38 No. 2, pp. 211-225, doi: [10.1108/EDI-09-2017-0182](https://doi.org/10.1108/EDI-09-2017-0182).
- Blair-Loy, M. and Cech, E.A. (2022), *Misconceiving Merit: Paradoxes of Excellence and Devotion in Academic Science and Engineering*, University of Chicago Press, Chicago, IL.
- Bonilla-Silva, E. and Peoples, C.E. (2022), "Historically white colleges and universities: the unbearable whiteness of (most) colleges and universities in America", *American Behavioral Scientist*, Vol. 66 No. 11, pp. 1490-1504, doi: [10.1177/00027642211066047](https://doi.org/10.1177/00027642211066047).
- Campbell, C.M. and O'Meara, K. (2013), "Faculty agency: departmental contexts that matter in faculty careers", *Research in Higher Education*, Vol. 55 No. 1, pp. 49-74, doi: [10.1007/s11162-013-9303-x](https://doi.org/10.1007/s11162-013-9303-x).
- Casad, B.J., Garasky, C.E., Jancetic, T.R., Brown, A.K., Franks, J.E. and Bach, C.R. (2022), "U.S. women faculty in the social sciences also face gender inequalities", *Frontiers in Psychology*, Vol. 13, 792756, doi: [10.3389/fpsyg.2022.792756](https://doi.org/10.3389/fpsyg.2022.792756), available at:
- Cate, L., Ward, L.W.M. and Ford, K.S. (2022), "Strategic ambiguity: how pre-tenure faculty negotiate the hidden rules of academia", *Innovative Higher Education*, Vol. 47 No. 5, pp. 795-812, doi: [10.1007/s10755-022-09604-x](https://doi.org/10.1007/s10755-022-09604-x).
- Cicchetti, D.V. (1994), "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology", *Psychological Assessment*, Vol. 6 No. 4, pp. 284-290, doi: [10.1037//1040-3590.6.4.284](https://doi.org/10.1037//1040-3590.6.4.284).
- Colby, G. and Fowler, C. (2020), *Data Snapshot: IPEDS Data on Full-Time Women Faculty and Faculty of Color*, American Association of University Professors, Washington, DC.
- Deci, E.L., Ryan, R.M., Gagné, M., Leone, D.R., Usunov, J. and Kornazheva, B.P. (2001), "Need satisfaction, motivation, and well-being in the work organizations of a former eastern bloc country: a cross-cultural study of self-determination", *Personality and Social Psychology Bulletin*, Vol. 27 No. 8, pp. 930-942, doi: [10.1177/0146167201278002](https://doi.org/10.1177/0146167201278002).
- Fox Tree, J.E. and Vaid, J. (2022), "Why so few, still? Challenges to attracting, advancing, and keeping women faculty of color in academia", *Frontiers in Sociology*, Vol. 6, 792198, doi: [10.3389/fsoc.2021.792198](https://doi.org/10.3389/fsoc.2021.792198), available at:
- Hardré, P. and Cox, M. (2009), "Evaluating faculty work: expectations and standards of faculty performance in research universities", *Research Papers in Education*, Vol. 24 No. 4, pp. 383-419, doi: [10.1080/02671520802348590](https://doi.org/10.1080/02671520802348590).
- Hill Collins, P. and Bilge, S. (2016), *Intersectionality*, Polity Press, Cambridge, GB.
- Holmes, M.H., Jackson, J.K. and Stoiko, R. (2016), "Departmental dialogues: facilitating positive academic climates to improve equity in STEM disciplines", *Innovative Higher Education*, Vol. 41 No. 5, pp. 381-394, doi: [10.1007/s10755-016-9358-7](https://doi.org/10.1007/s10755-016-9358-7).
- Hughes, B.E., Smith, J.L., Bruun, M., Shanahan, E.A., Rushing, S., Intemann, K., Handley, I.M., Belou, R., Stoop, C. and Sterman, L. (2022), "Department leaders as critical conduits for the advancement of gender equity programs", *Journal of Women and Gender in Higher Education*, Vol. 15 No. 1, pp. 41-64, doi: [10.1080/26379112.2022.2034122](https://doi.org/10.1080/26379112.2022.2034122).
- Jackson, J.K., Anderson, A.O., Dilks, L., Holmes, M.H., Kunkle, C.E., Nolan, J. and Latimer, M. (2023), *Engaging Faculty in Group-Level Change for Institutional Transformation: Disrupting Inequity and Building Inclusive Academic Departments*, Routledge, New York, NY.
- Järvinen, M. and Mik-Meyer, N. (2024), "Giving and receiving: gendered service work in academia", *Current Sociology*. doi: [10.1177/00113921231224754](https://doi.org/10.1177/00113921231224754).
- Kaiser, C.R., Major, B., Jurcevic, I., Dover, T.L., Brady, L.M. and Shapiro, J.R. (2013), "Presumed fair: ironic effects of organizational diversity structures", *Journal of Personality and Social Psychology*, Vol. 104 No. 3, pp. 504-519, doi: [10.1037/a0030838](https://doi.org/10.1037/a0030838).



- Krukowski, R.A., Jagsi, R. and Cardel, M.I. (2021), "Academic productivity differences by gender and child age in science, technology, engineering, mathematics, and medicine faculty during the COVID-19 pandemic", *Journal of Women's Health*, Vol. 30 No. 3, pp. 341-347, doi: [10.1089/jwh.2020.8710](https://doi.org/10.1089/jwh.2020.8710).
- Kuvaas, B. (2006), "Performance appraisal satisfaction and employee outcomes: mediating and moderating roles of work motivation", *The International Journal of Human Resource Management*, Vol. 17 No. 3, pp. 504-522, doi: [10.1080/09585190500521581](https://doi.org/10.1080/09585190500521581).
- Latimer, M., Jackson, K., Dilks, L., Nolan, J. and Tower, L. (2014), "Organizational change and gender equity in academia: using dialogical change to promote positive departmental climates", in *Gender Transformation in the Academy*, Emerald Group Publishing, Cambridge, UK, pp. 333-353.
- Milligan, S. (2016), "The flexibility paradox", *HR Magazine*, April 15, available at: <https://www.shrm.org/hr-today/news/hr-magazine/0416/pages/the-flexibility-paradox.aspx>
- Misra, J., Lundquist, J.H. and Templer, A. (2012), "Gender, work time, and care responsibilities among faculty", *Sociological Forum*, Vol. 27 No. 2, pp. 300-323, doi: [10.1111/j.1573-7861.2012.01319.x](https://doi.org/10.1111/j.1573-7861.2012.01319.x).
- Mitchneck, B., Smith, J.L. and Latimer, M. (2016), "A recipe for change: creating a more inclusive academy", *Science*, Vol. 352 No. 6282, pp. 148-149, doi: [10.1126/science.aad8493](https://doi.org/10.1126/science.aad8493).
- Niemann, Y.F., Gutierrez y Muhs, G. and Gonzales, C.G. (2020), *Presumed Incompetent II: Race, Class, Power, and Resistance of Women in Academia*, Utah State University Press, Logan, UT, available at: <https://upcolorado.com/utah-state-university-press/item/3794-presumed-incompetent-ii>
- O'Meara, K. (2005), "Encouraging multiple forms of scholarship in faculty reward systems: does it make a difference?", *Research in Higher Education*, Vol. 46 No. 5, pp. 479-510, doi: [10.1007/s11162-005-3362-6](https://doi.org/10.1007/s11162-005-3362-6).
- O'Meara, K., Eatman, T. and Petersen, S. (2015), "Advancing engaged scholarship in promotion and tenure: a roadmap and call for reform", Vol. 101 No. 3, Association of American Colleges and Universities, available at: <https://www.aacu.org/liberaleducation/2015/summer/o%27meara>
- O'Meara, K., Templeton, L., Culpepper, D. and White-Lewis, D. (2022), *Translating Equity-Minded Principles into Faculty Evaluation Reform*, American Council on Education, Washington, DC, available at: <https://www.acenet.edu/Documents/Equity-Minded-Faculty-Evaluation-Reform.pdf>
- Schein, E.H. and Schein, P.A. (2021), *Humble Inquiry: the Gentle Art of Asking Instead of Telling*, Penguin Random House, Manhattan, NY.
- Whitford, E. (2020), "College presidents' cabinets still far from gender parity", *Inside Higher Ed*, 15 April, available at: <https://www.insidehighered.com/news/2020/04/15/pay-and-seniority-gaps-persist-women-and-minority-administrators-higher-education>

**Corresponding author**

Jessi L. Smith can be contacted at: [jsmith20@uccs.edu](mailto:jsmith20@uccs.edu)

For instructions on how to order reprints of this article, please visit our website:

[www.emeraldgroupublishing.com/licensing/reprints.htm](http://www.emeraldgroupublishing.com/licensing/reprints.htm)

Or contact us for further details: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)