Why self-appraisals may not be best way to judge job performance - Harvard Gazette

Source: gnews-performance-review | **Published:** December 10, 2025 | **PMS Relevance:** PMS CORE | **Type:** Research Study | **Geography:** Global

Source URL: https://news.google.com/rss/articles/CBMisgFBVV95cUxOc2JhQ3lRZEFRTmxtRTdudUZfMHF

Executive Summary

This article reports on research co-authored by Iris Bohnet, Albert Pratt Professor of Business and Government at Harvard Kennedy School, examining how gender and race influence performance ratings at a multinational financial services company. The central argument is that sharing employee self-appraisals with managers prior to managerial review creates an 'anchoring' effect, where manager scores closely mirror self-ratings. The research finds that women and workers of color consistently rate themselves lower, with women of color assigning themselves the lowest self-ratings of any group. Managers, in turn, rated people of color more harshly than white employees, while showing a partial 'gender boost' that reduced — but did not eliminate — disadvantages for women. A system glitch that prevented managers from viewing self-appraisals in one year showed reduced correlation between self- and manager ratings, though racial and gender dynamics persisted, partly because managers appeared to reference prior-year self-evaluations. The authors conclude that demographic patterns in ratings are sufficiently correlated with social identity to raise concerns, even absent definitive proof of bias, and the company subsequently ceased sharing self-evaluations with managers prior to review.

Key Insights

1. Women and workers of color systematically assign themselves lower self-ratings, with women of color rating themselves the lowest of any demographic group studied.

2. Manager ratings exhibit a strong 'anchoring' effect when managers view employee self-appraisals beforehand — scores closely correlate with those self-ratings, embedding pre-existing self-evaluation gaps into official performance records.

3. Even when the anchoring mechanism was disrupted by a system glitch, racial and gender disparities in manager ratings persisted, partly because managers reverted to prior-year self-evaluations as a reference point.

Practical Takeaways

• Organizations that allow managers to view employee self-appraisals before submitting their own ratings may be inadvertently encoding demographic disparities into official performance outcomes.

• Data analysis of internal performance rating patterns — disaggregated by gender, race, and employee tenure — can serve as a diagnostic tool for identifying potential inequities in appraisal systems.

Critical Analysis

Strengths: The research draws on real organizational data from a multinational company, including a natural experiment created by a system glitch, lending observational rigor. The authors are transparent about the limits of their causal claims, explicitly stating that bias cannot be proven without objective performance benchmarks. The interview format allows nuanced qualification of findings. Limitations: The study relies on a single company's data, limiting generalizability. The binary treatment of gender and race is acknowledged as a constraint of global data collection rather than a methodological choice. The 'glitch year' natural experiment is suggestive but not a controlled intervention. The persistence of disparities even without self-appraisal sharing — attributed to managers consulting prior-year ratings — complicates the policy implication that simply withholding self-evaluations resolves inequity. Bias considerations: The article is co-authored by a prominent researcher with a stated focus on gender equity, which may frame findings in a direction consistent with pre-existing hypotheses. The company's decision to act on the research is presented without independent validation of outcomes.

Organizational Practice

The article describes a multinational financial services company that collects annual employee self-appraisals and shares them with managers prior to managerial review. The company also experienced a system glitch in one year that prevented self-appraisals from being shared in advance. Following the research analysis, the company discontinued the practice of sharing self-appraisals with managers before they submit their own ratings, and introduced regular internal data analysis to monitor demographic patterns in performance scores.

Strategic Implications

The findings point toward growing scrutiny of the sequencing and visibility of self-appraisal data within performance management workflows. The anchoring effect identified suggests that process design — specifically the order in which information is presented to evaluators — has measurable consequences for rating outcomes across demographic groups. The persistence of disparities even when anchoring is disrupted implies that structural inequity in performance systems may be multi-causal and resistant to single-point interventions. The use of internal data analytics as a diagnostic mechanism represents an emerging organizational practice for monitoring equity in performance outcomes. The trend toward higher-frequency, lower-stakes feedback cycles — quarterly or continuous — is noted as an area where evidence remains inconclusive.

Validation Status: passed
