

Gender and cultural bias in student evaluations: Why representation matters

Source: PLOS ONE | **Author:** Fan et al. | **PMS Relevance:** PMS RELATED | **Type:** Research Study | **Geography:** Asia-Pacific

Source URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0209749>

Executive Summary

This article addresses the presence of gender and cultural bias in student evaluations of teaching (SET) within higher education. The authors argue that SET scores — widely used in academic performance management and promotion decisions — are subject to systematic bias against female instructors and those from non-English speaking backgrounds. Drawing on 523,703 individual student surveys across five faculties at a large Australian public university (UNSW) over 2010–2016, covering 3,123 teachers and 2,392 unique courses, the study employs a cumulative ordinal regression model with random effects to control for course and teacher variation. Key findings include: female instructors from non-English speaking backgrounds face the strongest negative bias, with local male students in Science giving them odds of a higher score approximately 42% that of male English-speaking instructors; bias is substantially reduced in course evaluations (where teachers are not directly assessed), suggesting students evaluate the person rather than the teaching; and faculties with higher proportions of minority group teachers exhibit lower bias levels, suggesting a correlation between representation and reduced bias. The study concludes that the magnitude of these biases may render SET scores unreliable as measures of teaching effectiveness for performance management purposes.

Key Insights

1. Female instructors from non-English speaking backgrounds face the most severe bias in SET scores, with odds of receiving higher scores from local male students in Science approximately 42% those of male English-speaking instructors.
2. Bias largely disappears in course evaluations (where students assess the course rather than the teacher), suggesting the effect reflects judgements about the person rather than teaching quality.
3. Higher proportional representation of minority groups within a faculty is correlated (approximately $r=0.5$) with reduced bias in SET scores, indicating that workforce composition may influence evaluation outcomes.
4. The magnitude of gender and cultural bias in SET scores can exceed the effect of a measurable teaching effectiveness proxy (first-time course teaching), particularly in the Business faculty.
5. No significant difference in bias was detected between undergraduate and postgraduate students, suggesting the biases are culturally ingrained rather than context-specific to the university environment.

Practical Takeaways

- Institutions relying on SET scores for promotion and performance decisions are, according to this study, using a measure subject to statistically significant gender and cultural bias that may disadvantage women and non-English speaking background staff.
- The correlation between faculty-level representation of minority groups and reduced bias levels indicates that workforce diversity may function as a structural moderator of evaluation bias in academic settings.

Critical Analysis

Strengths: This is the largest institution-wide observational study of SET bias to date (523,703 surveys), providing substantially greater statistical power than prior case studies. The use of a mixed ordinal regression model with random effects for both course and teacher accounts for key confounds. The comparison between teaching and course evaluations provides a natural internal control that strengthens the bias interpretation. **Limitations:** The study is observational and cannot establish causality — association between teacher demographics and lower scores does not confirm the mechanism is bias. Approximately 24% of teacher cultural background data was imputed using country-of-birth proxies, introducing potential misclassification. Survey response rates of around 30% mean findings reflect only responding students, potentially introducing non-response bias that the university administration does not correct for when using these scores. The use of 'teaching the course for the first time' as a proxy for teaching effectiveness is acknowledged as imperfect. The study is single-institution and located in an Australian context with a high international student population, limiting generalisability to other national or institutional contexts. Funding by the university's own Deputy Vice-Chancellor Academic division is noted, though the authors declare no competing interests and funders had no role in analysis or publication decisions.

Organizational Practice

The university described in the study uses institution-wide Course and Teaching Evaluation and Improvement (CATEI) surveys as a routine component of performance management, including for promotion decisions. SET scores from voluntary student surveys (approximately 30% response rate) are reported to management and used as performance indicators. The university operates a mature data warehouse built on the Kimball method to integrate and analyse these results. At the time of the study, the university was described as proactively seeking to address conscious and unconscious bias in its evaluation processes.

Strategic Implications

The findings indicate that performance evaluation systems in academic institutions that rely on student evaluations may embed structural disadvantages for female and culturally diverse staff, with implications for promotion equity and workforce composition over time. The correlation between higher minority representation and lower observed bias suggests that diversity in workforce composition may have a moderating effect on evaluation outcomes, which has implications for how institutions frame diversity initiatives — not only as equity measures but as potential moderators of evaluation system distortion. The finding that biases are consistent across undergraduate and postgraduate levels implies that graduating students may carry these evaluation tendencies into professional workplaces, with downstream implications for performance culture in organisations that recruit university graduates. The discrepancy between teaching and course evaluation scores points to a measurement design consideration relevant to any organisation using peer or stakeholder ratings of individuals versus outputs.

Validation Status: passed

Generated by Peoplense (peoplense.com) on 2026-07-21 09:22 UTC