



# BEST PRACTICES IN REDUCING BIAS IN STUDENT EVALUATIONS OF TEACHING

In the following report, Hanover Research presents the results of a literature review of bias in student evaluations of faculty. The analysis presents an overview of the problem of student bias and outlines strategies for mitigating bias in evaluation design and administration.



# INTRODUCTION



Research increasingly indicates that student evaluations of teaching (SET) are affected by a variety of factors *not* related to instructor effectiveness. Most existing research on evaluation bias examines the impact of bias on women and people of color, but a multitude of factors – including an instructor’s sexuality, native language, age, and physical appearance – can lead students to judge faculty overly harshly ([Inside Higher Ed](#), [University of Dayton](#)). In particular, recent studies have shown:

- Internalized bias can penalize **female** faculty up to **0.4 points** on a 5-point scale ([American Political Science Association](#)).
- Internalized bias can penalize **Black** faculty up to **0.6 points** on a 5-point scale ([Journal of Negro Education](#)).
- Online student evaluations have become increasingly “abusive and bullying,” which researchers speculate could be related to similar behaviors in anonymous, online comments sections ([AAUP](#)).

As colleges and universities assess the inclusivity of their campus climate, they should also take steps to ensure that their methods for evaluating faculty performance are as free from bias as possible.

## KEY FINDINGS

To reduce or eliminate bias in student evaluations of teaching, higher education providers should consider the following practices:

- **Ask concrete questions** – General questions such as “How would you rank this teacher’s overall effectiveness” are most likely to be influenced by student biases. Instead, ask specific questions related to timelines, actions, and behaviors (e.g., “Did this instructor return graded work within two weeks?”).
- **Administer evaluations on paper and in class** – Research demonstrates that in-person evaluations net higher response rates and discourage derogatory or otherwise irrelevant comments.
- **Include information about implicit bias within evaluation forms** – Researchers at Iowa State University found that prompting students about unconscious bias positively impacted reviews for female professors. This strategy is a simple, low-investment method to address bias in evaluation surveys.

# SET ADMINISTRATION STRATEGIES

## COMPLETE EVALUATIONS IN CLASS

Response rates for evaluations administered online and outside of class are considerably lower than those administered via paper and in-class settings. This may also skew responses towards students with the most positive and negative experiences ([AAUP](#)).

An American Association of University Professors faculty survey found that response rates for online, outside-of-class reviews ranged from 20 to 40 percent, compared to approximately 80 percent for paper, in-class evaluations. In addition, **paper, in-class evaluations appear to limit derogatory comments**, as they increase feelings of responsibility while maintaining confidentiality ([AAUP](#)).



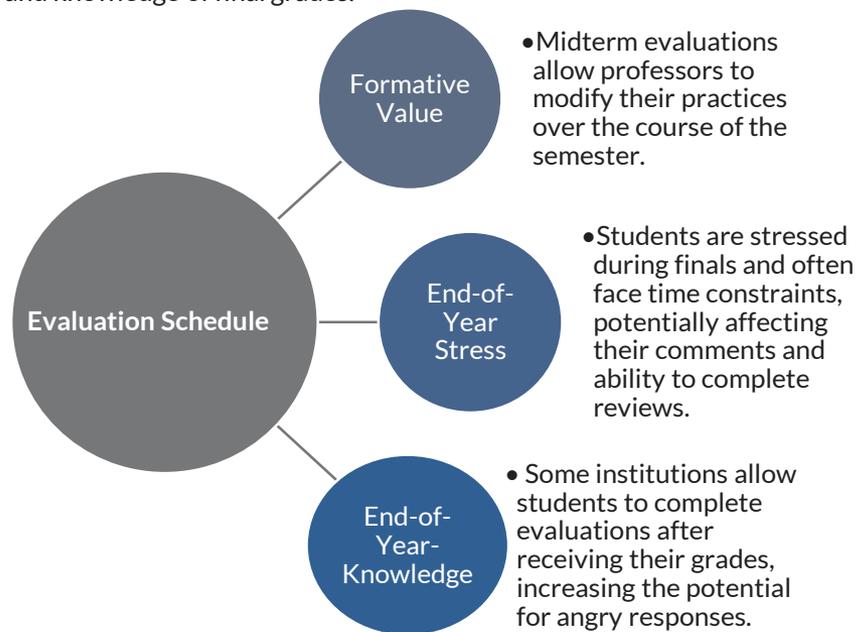
**Online evaluations have taken a harsher tone on matters** that do not relate to teaching (e.g., appearance). This trend particularly **affects women**, who are more likely to be described in terms of physical appearance or personal characteristics than professional ability ([Harvard](#)).

Information regarding individual institutions' SET providers is limited, but schools should typically avoid large providers that develop unreliable, cross-field benchmarks (e.g., Educational Testing Service; [Inside Higher Ed](#)). However, IDEA, another SET vendor, markets instruments that focus on concrete behaviors to help reduce bias ([Inside Higher Ed](#)).



## USE MIDTERM EVALUATIONS

Indiana University – Purdue University Indianapolis ([IUPUI](#)) recommends midterm evaluations as an alternative performance measure, and [AAUP](#) notes that end-of-term evaluations can be swayed by students' stress level and knowledge of final grades.



## UNIVERSITY OF OREGON: MIDWAY AND END-OF-COURSE EVALUATIONS

The University of Oregon (UO) conducts course evaluations [twice per term](#). The Midway Student Experience Survey is administered a few weeks into the term. Results are only available to instructors, enabling them to make adjustments based on student feedback. The End-of-Course Student Experience Survey is administered at the end of each term. Results are available to instructors, department heads, deans, and personnel and awards committees.

# SET DESIGN STRATEGIES

## ASK ABOUT CONCRETE BEHAVIORS

Ambiguous questions about overall ratings increase the possibility that implicit bias will affect students' answers ([IUPUI](#)). For instance, a review of data from RateMyProfessors.com found that students typically rate female instructors as less helpful than male instructors ([Assessment & Evaluation in Higher Education](#)). Researcher and assistant professor at the University of Louisville W. Carson Byrd posits that this finding may relate to perceptions of women as caregivers, increasing students' expectations of female instructors' accessibility and flexibility ([Inside Higher Ed](#)). The table below provides examples of how to make common student evaluation questions more concrete. Concrete questions should specify timeframes, specific behaviors, interactions, and the like.

Ambiguous	Concrete
Did your instructor return assignments promptly?	Did your instructor return graded assignment within two weeks?
Was your instructor accessible to students?	Was your instructor always in their office during office hours?

There are also types of questions that should be avoided entirely. Both [IUPUI](#) and [UC Berkeley](#) warn about “**global agreement**” statements and questions about overall effectiveness or course value. Examples include:



- Overall, how do you rate this instructor?
- Overall, how do you rate this course?
- What is the overall rating of the instructor's teaching effectiveness?

## ASK ABOUT STUDENT OBJECTIVES

[Guidelines from IUPUI](#) recommend asking questions about whether students felt their learning objectives were accomplished, whether they felt capable of success, learning methods they observed, and students' motivation in taking the course. These questions force more objective answers and provide evaluators with more context on the student.



Asking for observations about teaching methods is an example of asking questions about concrete behaviors. It forces students to think more deeply about the class, rather than citing “personality differences,” which can often hide bias ([Inside Higher Ed](#)).



Professors typically receive better ratings from students who are interested in the course topic (e.g., relates to their major or is an elective, compared to a required or general education course; [University of Dayton](#)).



Difficult, entry-level classes are particularly problematic. The director of Northwestern's Communication and Legal Reasoning program found evaluations for the first-year analysis, writing, and research course significantly disadvantaged their predominantly female faculty ([Chronicle of Higher Education](#)).

# SET DESIGN STRATEGIES

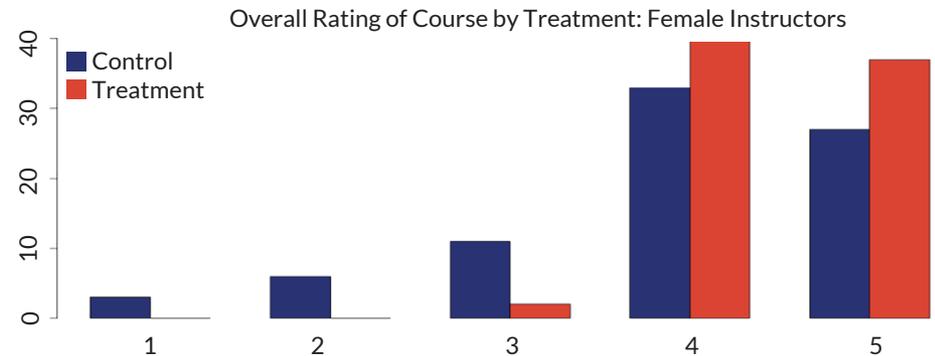
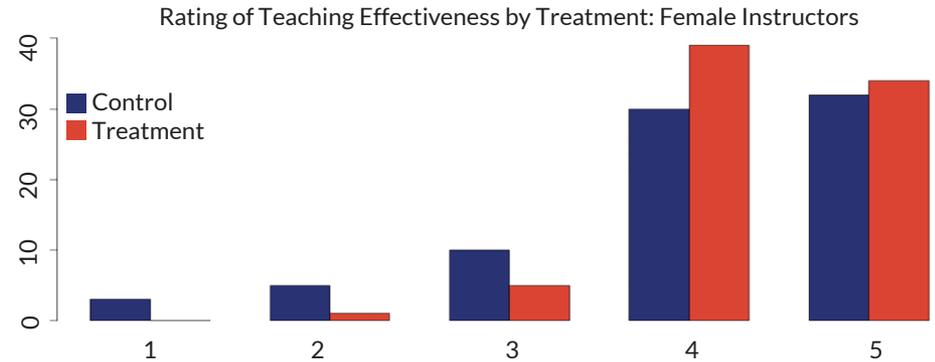
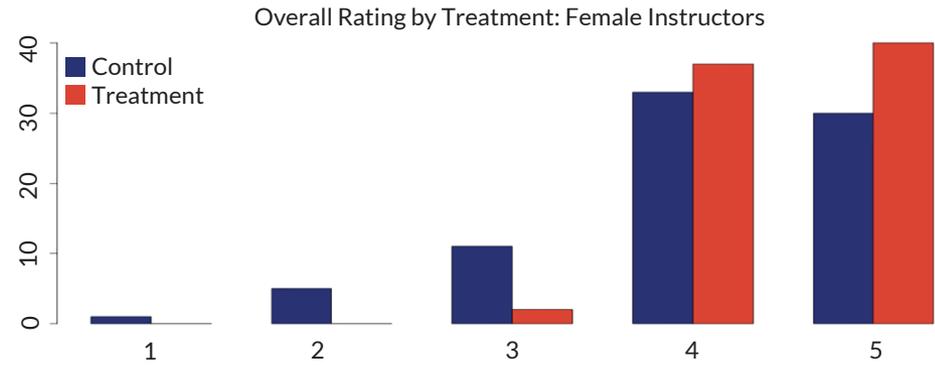
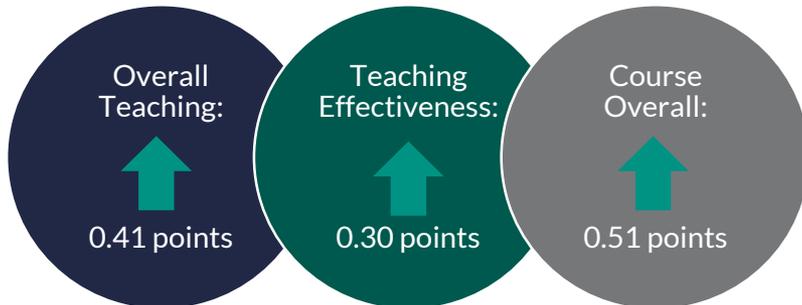
## INCLUDE INFORMATIVE TEXT

Researchers at Iowa State University found that adding text reminding students of the effects of unconscious bias positively impacted reviews of female professors (without impacting reviews of male professors). Students in two sections of two different courses (biology and politics) were randomly assigned to a control group (without additional text) or treatment group (with additional text). One section of each course was taught by a female professor, the other by a male professor. **The results of the study showed that students who received information about implicit bias:**



- Rated overall teaching an average of **0.41 points higher** for female professors
- Rated teaching effectiveness of female professors an average of **0.30 points higher**
- Rated the overall course taught by female professors an average of **0.51 points higher**

While the authors recommend using this strategy, there is a possibility that its widespread implementation could lessen the impact as students become used to seeing the language on bias regularly ([PLOS ONE](#)).



# PROFILE: IOWA STATE UNIVERSITY

## IOWA STATE UNIVERSITY OVERVIEW

[Iowa State University](#), which experimented with adding text reminding students of implicit biases to its student evaluations, provides the following guidelines for student evaluation instruments:

Include **open- and close-ended** questions

Include intentional measures of both general instructor **attributes** (e.g. enthusiasm or effectiveness) and specific instructor **behaviors** (e.g. listening, providing feedback)

Use **consistent scales** (e.g., five-points, same direction, 1 = low, 5 = high) and a no-opinion option

The evaluations should produce **useful feedback** to instructors that can inform their teaching

Students should be able to complete them thoughtfully within **10 to 15 minutes**

Source: Adapted directly from [Iowa State University](#)

## IMPLICIT BIAS LANGUAGE

Researchers at Iowa State found that adding a reminder highlighting the usage of evaluations and that bias most often negatively impacts female faculty and faculty members of color reduced unconscious bias in SETs. An example of such text is shown below.

*“Student evaluations of teaching play an important role in the review of faculty. Your opinions influence the review of instructors that takes place every year. Iowa State University recognizes that student evaluations of teaching are often influenced by students’ **unconscious and unintentional** biases about the race and gender of the instructor. Women and instructors of color are systematically rated lower in their teaching evaluations than white men, even when there are no actual differences in the instruction or in what students have learned.*

*As you fill out the course evaluation please keep this in mind and make an effort to resist stereotypes about professors. Focus on your opinions about the content of the course (the assignments, the textbook, the in-class material) and not unrelated matters (the instructor’s appearance).”*

# SET DESIGN STRATEGIES

## ALLOW DEPARTMENT FACULTY TO DESIGN AND ASSESS EVALUATIONS

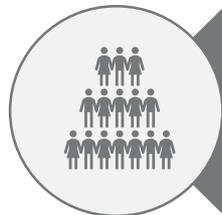
Faculty are closer to course goals, student ability, and promoted teaching styles than administrators. In addition, the content and goals of departments can vary so broadly that institution-wide uniformity “generate(s) meaningless numerical comparisons that invite misuse” ([AAUP](#)).



Controversial course topics can sway student opinions, which should be considered in evaluations. ([Inside Higher Ed](#)).



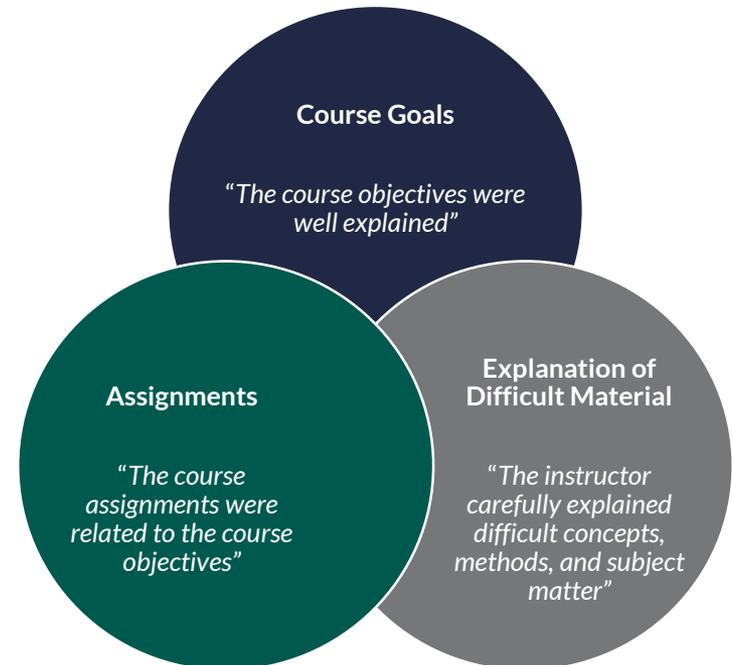
On average, faculty in arts and humanities receive higher ratings than those in social sciences and math; the cause is not clear, but data indicate that different fields may be more difficult to teach and may use different instructional styles ([University of Dayton](#)).



Teaching large lectures requires a different style than smaller seminars; using the same instrument for evaluations of these courses is not valid ([AAUP](#)).

## EMPHASIZE LEARNING AS A SHARED RESPONSIBILITY

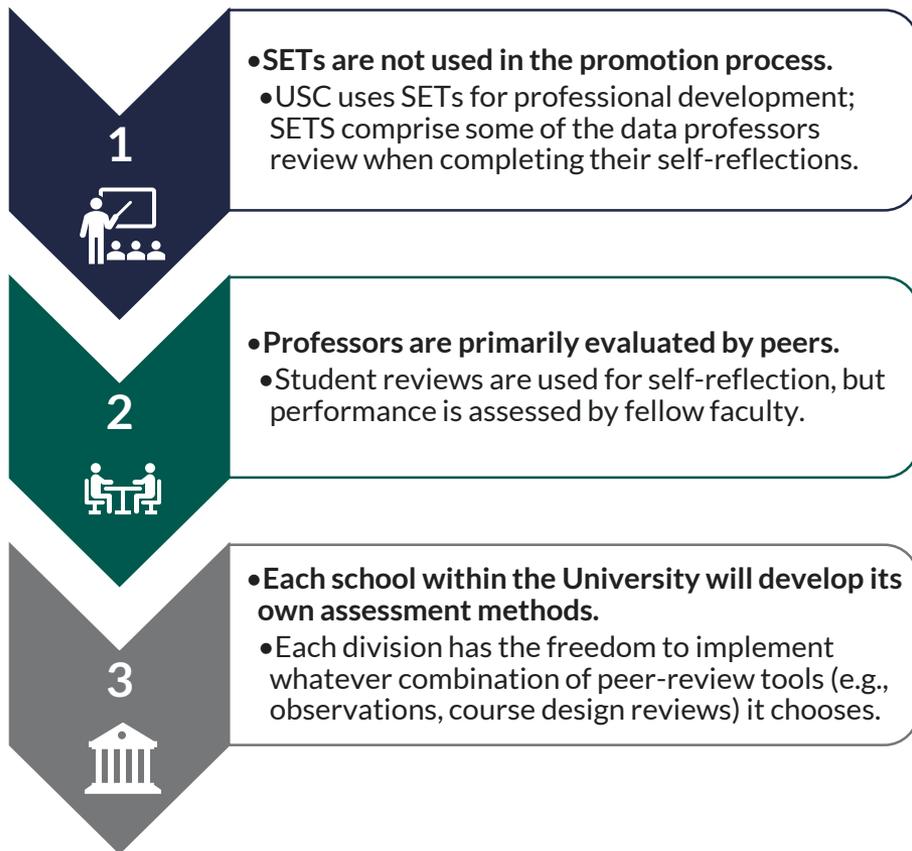
Several studies ([AAUP](#), [Assessment and Evaluation in Higher Education](#), and [UC Berkeley](#)) demonstrate that students prefer easy courses and rank professors based on their anticipated grade. The University of Southern California revised its evaluations to ask questions about course goals, the relevancy of assignments, and explanations of difficult concepts, rather than solely how students perceived course difficulty or value. ([IUPUI](#)).



# PROFILE: UNIVERSITY OF SOUTHERN CALIFORNIA

## USC OVERVIEW

In 2018, the [University of Southern California](#) redesigned its student evaluation instrument and professional review process to better account for student bias. Changes included:



## OLD USC SURVEY INSTRUMENT

1. How would you rate the instructor's effectiveness on the following items?
  1. Clearly articulated course goals
  2. Organized course to achieve those goals
  3. Carefully explained difficult concepts, methods, and subject matter
  4. Encouraged students to participate in their learning (e.g., through discussion, projects, study groups and other appropriate activities)
  5. Was accessible to students (e.g., during office hours, before and after class, etc.)
  6. Evaluated student work in fair and appropriate ways
  7. Was enthusiastic about communicating the subject matter
  8. Stimulated student interest in the subject matter
  9. Presented subject matter in ways that were academically challenging
  10. Provided students a valuable learning experience
2. Overall, how would you rate this instructor?
3. Overall, How would you rate this course?
4. What were this instructor's main strengths?
5. How might this instructor improve his or her teaching effectiveness?
6. Additional comments?

# PROFILE: UNIVERSITY OF SOUTHERN CALIFORNIA

## NEW USC SURVEY INSTRUMENT

### COURSE DESIGN

1. *The course objectives were well explained.*
2. *The course assignments were related to the course objectives.*
3. *I understood what was expected of me in this course.*

### INSTRUCTIONAL PRACTICES

1. *The instructor carefully explained difficult concepts, methods, and subject matter.*
2. *The instructor encouraged questioning and discussion of course topics from the students.*
3. *The instructor encouraged me to do my best work.*

### INCLUSION PRACTICES

1. *The course materials included diverse perspectives OR application to diverse populations.*
2. *The instructor used a variety of teaching approaches to meet the needs of all students.*
3. *The instructor was receptive to the expression of diverse student viewpoints.*
4. *The instructor demonstrated sensitivity to students' needs and diverse life experiences.*

### ASSESSMENT PRACTICES

1. *The assessments/assignments reflected what was covered in the course.*
2. *The grades I have received thus far reflect the QUALITY of my performance in the course.*
3. *The criteria for good performance on the assignments or assessments were clearly communicated.*

### COURSE IMPACT

1. *I learned perspectives, principles, or practices from this course that I expect to apply to new situations.*
2. *This course challenged me to think critically and communicate clearly about*

*the subject.*

3. *This course provided me with information that may be directly applicable to career or academic goals.*

*If you disagreed with one of the previous items on Course Design, Instructional Practices, Assessment Practices, or Course Impact, please describe a change that would improve that aspect of the course.*

### STUDENT ENGAGEMENT

1. *Approximately how many hours did you spend doing homework for this course?*
  1. *1-3 hours per week*
  2. *4-7 hours per week*
  3. *Greater than 7 hours per week*
2. *Approximately how many times did you interact with the instructor outside of class (e.g., via email, office hours)?*
  1. *1-3 times during the semester*
  2. *4-7 times during the semester*
  3. *Greater than 7 times during the semester*
  4. *I did not interact with the instructor outside of class*
3. *In what ways have you participated in your learning for this course?*
  1. *I studied with students in the class*
  2. *I sought the assistance of the instructor during class*
  3. *I sought assistance outside of class*
  4. *I engaged in discussions in class*
  5. *I used electronic resources*
  6. *I studied alone*
  7. *I kept current with course readings*
  8. *I kept current with course assignments*
  9. *Other, please describe*
4. *Please describe the most valuable aspect(s) of this course.*
5. *Please describe the least valuable aspect(s) of this course.*
6. *Is there additional information or feedback that you would like to share with the course instructor?*



HANOVER  
RESEARCH

