STUDENT EVALUATIONS AND THE PROBLEM OF IMPLICIT BIAS

ROGER W. REINSCH, SONIA M. GOLTZ, AMY B. HIETAPELTO*

"It is easy to believe that there is more going on in people's minds than they say; it is not easy to believe that there is more going on in my mind than I say."[1]

INTRODUCTION

This article addresses the implicit bias problems inherent in using student evaluations when making employment decisions concerning university faculty members. Research indicates that student evaluations contain implicit bias regarding race, gender, and a variety of other protected categories. We begin by looking at the current use, purpose and structure of student evaluations. We then explore what implicit bias is and the research that demonstrates that most of us have some sort of implicit bias. Once the concept of implicit bias is explained, we examine the research that indicates there is implicit bias in student evaluations. We then discuss the law and implicit bias generally, followed by specific legal issues that are raised. Next, we examine recent trends at some universities which have recognized and begun to address the problems with student evaluations. Finally, we offer recommendations as to how to evaluate faculty members' teaching using alternative methods.

I. Use and Purpose of Student Evaluations of Teaching:

Student Evaluations of Teaching[2] are recognized as a common performance measure used by universities to make employment decisions with regard to faculty, as Emery and colleagues noted: "A current practice among colleges and universities in the USA is for the administration to use a student evaluation instrument of teaching effectiveness as part of the faculty member's performance evaluation."[3]

---

* Roger W. Reinsch, Professor, Labovitz School of Business and Economics, Juris Doctorate from the University of Missouri-Columbia. Sonia M. Goltz, Professor, School of Business and Economics, Michigan Technological University, Ph.D. Industrial/Organizational Psychology, Purdue University. Amy B. Hietapelto, Professor and Dean, Labovitz School of Business and Economics, Ph.D. in Business Administration, University of Minnesota-Twin Cities.

1 Brian A Nosek, Carlee Beth Hawkins and Rebecca S. Frazier, *Implicit Social Cognition: From Measures to Mechanisms*, 15 Trends Cogn. Sci. 152 (2011).

2 SETs hereafter.

3 Charles R. Emery, Tracy R. Kramer and Robert G. Tian, *Return to Academic Standards: A Critique of Student Evaluations of Teaching Effectiveness*, 11 Quality Assurance in Education 37, 37 (2003).

Faculty at most colleges and universities today in the United States are subject to summative student evaluations.[4] These summative student evaluations are used to make several employment decisions, such as when determining pay increases, tenure and promotion.[5]

Summative student evaluations of teaching regularly use numerical scores to assess whether a faculty member is a good teacher.[6] The intended purpose of summative evaluations is to provide information to administrators about the faculty member's teaching ability. However, student ratings may represent essentially little more than opinions, raising the issue of potential student bias as Hornstein noted: "The validity of anonymous students' evaluations rests on the assumption that, by attending lectures, students observe the ability of the instructors, and that they report it truthfully"[7]. Institutions are typically not unaware of the likelihood of bias coloring evaluations, but use them anyway largely because of their convenience, as noted by Flaherty: "While some institutions have acknowledged the biases inherent in SETs, many cling to them as a primary teaching evaluation tool because they're easy -- almost irresistibly so. That is, it takes a few minutes to look at professors' student ratings on, say, a 1-5 scale, and label them strong or weak teachers. It takes hours to visit their classrooms and read over their syllabi to get a more nuanced, and ultimately more accurate, picture."[8] Therefore, many administrators seem willing to discount or overlook the possibility of bias so they can continue to rely on SETs.

As stated, the evaluations typically use a Likert scale anchored with numbers, often from 1-5. These numbers are often associated with verbal anchors—for example, five usually means a high rating. Student evaluations are then compiled, producing a mean score for each question, and finally an overall mean score for that faculty member in that class. Those making employment decisions generally will rely mostly on the overall mean scores.[9] At most institutions, although a mean

---

4 *See generally*, S. Surbhi, *Difference Between Formative and Summative Assessment*, (July 1, 2017), https://keydifferences.com/difference-between-formative-and-summative-assessment.html, Summative assessment is an assessment of learning that is normally done at the end of the semester. Essentially students are asked to evaluate their learning in a course once it is over. In effect, it is their opinion of what they learned. In contrast, formative assessment is done on an on-going basis during the semester to assess learning at each of those intervals. Its purpose is to assess whether the pedagogy used is conducive for learning. It provides information that may be used to change the pedagogy during the semester. Whereas, summative assessment does not provide an opportunity to change anything during that semester.

5 *See*, Henry A. Hornstein, *Student Evaluations of Teaching are an Inadequate Assessment Tool for Evaluating Faculty Performance*, 4 Cogent Education 1 (2017).

6 *See*, John W. Lawrence, *Student Evaluations of Teaching are Not Valid*, AAUP, (May-June 2018), https://www.aaup.org/article/student-evaluations-teaching-are-not-valid#.XFCj6fZFyUk.

7 Hornstein, *supra* note 5 at 3.

8 Coleen Flaherty, *Teaching Eval Shake-Up*, Inside Higher Ed (May 22, 2018) at https://www.insidehighered.com/news/2018/05/22/most-institutions-say-they-value-teaching-how-they-assess-it-tells-different-story.

9 For example, at one of the authors' institutions this is used: Level 3 – truly meritorious Exceeds expectations. Faculty member (1) is a highly effective teacher, (2) engages students in a variety of meaningful ways, and (3) has student-based teaching evaluations that consistently include

of 3 is usually designated as "acceptable" with the scale's verbal anchor, being in the 3 range is considered by administrators not to be very good and, in fact, may be seen as problematic. The expectation is that everyone will be at 4 or above. This expectation is a false one and could be viewed as a manifestation of the Garrison Keillor syndrome, namely that "everyone is above average."[10]

Because the expectation is that everyone should be rated above average, the presence of implicit bias is even more concerning. Bias is likely to lower a particular faculty member's mean score while at the same time raising another faculty member's score. This makes it difficult for certain groups of people—usually underrepresented group members—to achieve "above average" ratings while making it easier for members of majority groups to do so. Therefore, if implicit bias involves any of the protected categories under the law and evaluations are used to make employment decisions, then those employment decisions are based on some factors that are discriminatory and therefore illegal. For instance, a lower mean could result in the faculty member receiving lower merit increases or not getting promoted. This would be discriminatory under The Equal Pay Act[11] and Title VII of the Civil Rights Act of 1964 as amended Civil Rights Act of 1991.[12]

For these reasons it is important to look at the types of biases that might be in these evaluations. The issue of likely bias should not be dismissed because it is inconvenient or a challenge to come up with alternative unbiased measures of performance; we believe instead that it should be treated as a critical issue because student evaluations of faculty are used frequently and in a variety of employment related decisions: SETs play a role in the hiring process, tenure decisions, promotion decisions, salary decisions, and other benefits such as faculty awards. This is not a new concern and has been recognized in the education literature, such as by Basow and Martin, who noted: "The question of whether student evaluations can be biased is a critical one for those using them, whether for formative or summative purposes."[13] Therefore, we will look at some of the research on implicit bias.

---

a significant number of item scores ≥ 5 (on a 6-point scale).

Level 2 – solid and sound Meets expectations. Faculty member (1) is an effective teacher, (2) engages students in meaningful ways, and (3) has student-based teaching evaluation scores that are average (not significantly high or low). (3.6-4.9)

Level 1 – substandard does not meet expectations. Faculty member (1) needs improvement in the area of teaching effectiveness, (2) does not demonstrate meaningful student engagement, and (3) has student-based teaching evaluations that consistently include a significant number of item scores ≤ 3.5 (6-point scale).

10 https://www.brainyquote.com/quotes/garrison_keillor_137097.

11 Equal Pay Act of 1963, 29 U.S. Code Chapter 8 § 206(d).

12 Civil Rights Act of 1964, Pub. L. No. 88-352, 78 Stat. 241 (codified as amended in scattered sections of 2 U.S.C., 28 U.S.C., and 42 U.S.C.).

13 Susan A. Basow and Julie L. Martin, *Bias in Student Evaluations* at 39. In M. E. Kite (Ed.), *Effective Evaluation of Teaching: A Guide for Faculty and Administrators*. Retrieved from the Society for the Teaching of Psychology, at http://teachpsych.org/ebooks/evals2012/index.php; for a definition of summative and formative evaluations *see supra*, note 4.

II. Implicit Bias: What is it?

Humans perceive other people's behavior through filters that are socially conditioned. None of us sees the world through neutral, objective lenses. Instead, our minds classify individuals according to race, gender, age, and other socially salient categories with dizzying speed.[14] Studying these social attitudes can be tricky, however; impression management, for example, can influence self-reports of social attitudes that are frowned on by society.[15] Also, individuals often have attitudes of which they are not fully aware. Therefore, in the 1980s, psychologists began to use indirect measures of attitudes that bypass conscious awareness, such as by relying on response latency[16] in order to better ascertain underlying mental processes.[17] In other words, these "implicit measures" do not require individuals to be aware of their attitudes.[18] Implicit measures are now widely used in personality and social psychology with about 20 implicit measurement methods having been developed.[19]

These measures are responsible for much of what we now know about implicit social cognition, a term Greenwald and Banaji introduced describing cognitive processes that occur outside of conscious awareness or control in relation to social psychological constructs - attitudes, stereotypes, and self-concepts.[20] There appear to be two distinct levels of social cognition. Much of human cognition that influences judgement and action seems to occur outside of conscious awareness or conscious control.[21] "At the lower level there are fast, relatively inflexible routines that are largely automatic and implicit and may occur without awareness. At the higher level there are slow, flexible routines that are explicit and require the expenditure of mental effort."[22] These two levels of decision making have been referred to as "System 1 and System 2" or "Fast and Slow" thinking.[23] "System I

---

14 *See,* Nalini Ambady, Frank J. Bernieri & Jennifer A. Richeson, *Toward a Histology of Social Behavior: Judgmental Accuracy from Thin Slices of the Behavioral Stream*, 32 Advances in Experimental Social Psychology 201, 247 (2000)

15 See, *e.g.* Russell H. Fazio & Michael A. Olson, *Implicit Measures in Social Cognition Research: Their Meaning and Use*, 54 Annual Review of Psychology 297, 300-303 (2003).

16 Fazio & Olson, *Id*. at 298-299

17 *Generally, see*, R. Duncan Luce, Response Times: Their Role in Inferring Elementary Mental Organization, New York, NY: Oxford University Press (1986).

18 Brian A. Nosek & Anthony G. Greenwald, *(Part of) The Case for a Pragmatic Approach to Validity: Comment on DeHouwer, Teige-Mocigemba, Spruyt, and Moors,* 135 Psychol Bull. 373, 374 (2009).

19 *See*, Judith Znanewitz, Lisa Braun, David Hensel, Claudia Fatapie Altobelli, & Fabian Hattke, *A Critical Comparison of Selected implicit Measurement Methods*, 11 Journal of Neuroscience, Psychology, and Economics 249 (2018).

20 *See*, Anthony G. Greenwald &Mahzarin R. Banaji, *Implicit Social Cognition: Attitudes, Self-esteem, and Stereotypes*, 102 Psychol Rev. 4, 4-5 (1995).

21 *See*, Brian A. Nosek, Carlee Beth Hawkins and Rebecca S. Frazier, *Implicit Social Cognition: From Measures to Mechanisms*, 15 Trends in Cognitive Science 152 (2011).

22 Chris D. Frith and Uta Frith, *Implicit and Explicit Processes in Social Cognition*, 60 Neuron 503, 504 (2008).

23 Kahneman, D. 2011.Thinking, Fast and Slow. Publishers, Farrar, Straus, and Giroux. Kindle Edition.

is rapid, intuitive, and error-prone; System II is more deliberative, calculative, slower, and often more likely to be error-free."[24]

The Implicit Association Test[25] (IAT) is one of the more well-known implicit measures found in social psychology and is designed to assess implicit bias, which refers to a preference for or against something that is outside of awareness.[26] The IAT measures the association between categories such as old and young, black and white, female and male, and value attributes such as pleasant and unpleasant or good and bad.[27] Meta-analyses indicate that, in contrast to explicit measures of stereotypes, implicit measures like the IAT are predictive across target groups and also predict equally well across behaviors that vary in controllability and conscious awareness.[28] Furthermore, they have been found to be more predictive of behavior than self-reported attitudes for socially sensitive topics.[29] In fact, in several areas, including law, healthcare, and business, implicit measures are used to answer questions of why inequities are still present even though expressed attitudes are often neutral.[30]

Implicit bias is based on stereotypes that are learned as part of growing up in a certain culture and/or environment. A stereotype is construct, in other words, a set of thoughts and beliefs, that contains a theory about a social group and influences social behavior.[31] For example, gender stereotypes are very prescriptive—the characteristics ascribed to women and men tend to set up expectations of behaviors from those groups.[32] This is true as well for other cultural stereotypes, such as race.[33] Stereotypes are based on a kernel of truth about differences between groups

---

24 Christine Jolls and Cass R. Sunstein, *The Law of Implicit Bias*, 94 Calif. L. Rev. 969, 974 (2006) (citations omitted).

25 *See*, http://implicit.harvard.edu.

26 *See, e.g.* John F. Dovidio, *On the Nature of Contemporary Prejudice: The Third Wave*, 57 Journal of Social Issues, 829, 838-839 (2001); Anthony G. Greenwald, T. Andrew Poehlman, Eric Luis Uhlmann & Mahzarin R. Banaji, *Understanding and Using the Implicit Association Test: III. Meta-analysis of Predictive Validity*, 97 J Pers Soc Psychol. 17, 18 (2009).

27 *See*, Anthony G. Greenwald, Debbie E. McGhee & Jordan L. K. Schwartz, *Measuring Individual Differences in implicit Cognition: The Implicit Association Test*, 74 Journal of Personality and Social Psychology 1464, 1465 (1998). http://doi.org/10.1037//0022-3514.74.6.1464

28 Benedek Kurdi, Allison E. Seitchik, Jordan R. Axt, Timothy J. Carroll, Arpi Karapetyan, Neela Kaushik, Anthony J. Greenwald and Mahzarin R. Banaji, (2018, June 20). *Relationship Between the Implicit Association Test and Intergroup Behavior: A Meta-analysis*, 10 American Psychologist 1047 (2018).

29 Greenwald et al, *supra* note 27 at 1476.

30 John T. Josta, Laurie A. Rudman, Irene V. Blair, Dana R. Carney, Nilanjana Dasgupta, Jack Glaser & Curtis D. Harding, *The Existence of Implicit Bias is Beyond Reasonable Doubt: A Refutation of Ideological and Methodological Objections and Executive Summary of Ten Studies That No Manager Should Ignore*, 29 Research in Organizational Behavior 39, 46-53 (2009).

31 Richard D. Ashmore and Frances K. Del Boca, *Sex Stereotypes and Implicit Personality Theory: Toward a Cognitive-Social Psychological Conceptualization*, 5 Sex Roles 219, 228 -229 (citations omitted) (1979).

32 Deborah A. Prentice, & Erica Carranza, *What Women and Men Should Be, Shouldn't Be, Are Allowed to Be, and Don't Have to Be: The Contents of Prescriptive Gender Stereotypes*, 26 Psychology of Women Quarterly 269 (2002).

33 Susan A. Basow and Julie L. Martin, *Bias in Student Evaluations* 40-49 (Mary E. Kite, ed., *Effective*

but beliefs about individuals in those groups then tend to be distorted toward the representative types rather than reflective of the fact that individuals typically fall along a normal distribution on every dimension.[34]

In effect, stereotypes make life simpler for the individual doing the stereotyping because the individual does not have to deal with the identifying complexities of another. Stereotypes serve as a shortcut or, in more academic terms, as a decision heuristic. In fact, stereotypes are thought to be a type of "representativeness heuristic," which is essentially an assessment of a probability that an individual will have a certain characteristic.[35] Decision heuristics such as stereotypes are used particularly when there is a lack of information about a situation or person and when there is a lack of time to obtain the needed information.[36] Decision heuristics are often useful,[37] but of course, there is the inherent risk they will be inaccurate. This is often the case with stereotypes. Stereotypes may describe generalities across large groups of people based on historical circumstances but will not describe everyone accurately and will contain assumptions that are likely to be inaccurate.[38] This inaccuracy is largely a function of the fact that implicit bias tends to be triggered rapidly with little deliberation, as noted by Jolls and Sunstein: "We believe that the problem of implicit bias is best understood in light of existing analyses of System I processes. Implicit bias is largely automatic; the characteristic in question (skin color, age, sexual orientation) operates so quickly, in the relevant tests, that people have no time to deliberate."[39]

These implicit biases affect people's responses towards others, which then can adversely impact those individuals. Social psychologists have documented how a rater's perception of and reaction to another person can be affected by bias, either consciously or unconsciously, explaining behaviors such as the backlash towards agentic women in hiring decisions.[40] Implicit bias has been discussed as a factor significantly affecting various outcomes for individuals, ranging from work experiences to psychological and physical health, and has been found to be as

---

Evaluation of Teaching: A Guide for Faculty and Administrators, Society for Teaching of Psychology) (2012).

34 Pedro Bordalo, Katherine Coffman, Nicola Gennaioli & Andrei Shleifer, *Stereotypes*, 131 The Quarterly Journal of Economics 1753 (2016).

35 Bordalo, et al *Id*. at 1753.

36 *See, e.g*., Gerd Gigerenzer & Wolfgang Gaissmaier, *Heuristic Decision Making*, 62 Annual Review of Psychology, 451, 452-453 (2011).

37 Gigerenzer et al *Id*. at 473.

38 Generally, *see*, Jordan Carpenter, Daniel Preotiuc-Pietro, Lucie Flekova, Salvatore Giorgi, Courtney Hagan, Margaret L. Kern, Anneke E.K. Buffone, Lyle Ungar & Martin E. P. Seligman, *Real Men Don't Say "Cute": Using Automatic Language Analysis to Isolate Inaccurate Aspects of Stereotypes*, 8 Social Psychological and Personality Science 310 (2017).

39 Christine Jolls and Cass R. Sunstein, *The Law of Implicit Bias*, 94 Calif. L. Rev. 969, 975 (2006).

40 *See, e.g*., Monica Biernat, *Toward a Broader View of Social Stereotyping*, 58 American Psychologist 1019 (2003); Alice H. Eagly and Steven J. Karau, *Role Congruity Theory of Prejudice Toward Female Leaders*, 109 Psychological Review 573 (2002); Julie E. Phelan, Corinne A. Moss-Racusin & Laurie A. Rudman, *Competent Yet Out in the Cold: Shifting Criteria for Hiring Reflect Backlash Toward Agentic Women*, 32 Psychology of Women Quarterly 406 (2008).

damaging as overt discrimination.[41] Implicit bias can even affect whether a person lives or dies. The nature of individuals' interactions with health care professionals, for example, appears to be affected by implicit bias,[42] as is whether police use additional force during interventions.[43] The latter research has been used to account for the greater incidence of police shootings with certain racial groups.[44] For example, a study by Joshua Correll, et al, found that unconscious race bias played a large factor in an experiment when participants played a game in which the researchers systematically varied the race of a series of men who appeared on the computer screen.[45] The participants were instructed to shoot men holding guns and not to shoot men holding something innocent such as a wallet. Results were that the players were significantly more likely to shoot blacks holding innocent objects versus whites holding those same objects.[46] Collectively, then, research on implicit social cognition provides "incontrovertible evidence that thoughts, feelings, and actions are shaped by factors residing largely outside conscious awareness, control, and intention."[47] Despite this evidence, addressing implicit bias is not easy. Implicit attitudes are rooted in habitual responses and therefore are persistent and more difficult to alter than are explicit ones.[48] A major result of implicit bias towards certain groups of people is that over time, even seemingly minor behaviors accumulate and can have substantial impact. For example, implicit bias results in a tendency for women to be consistently underrated [49] and for women's work to be devalued.[50] Over time, this results in a large advantage for

---

41 *See, e.g.*, Kristen P. Jones, Chad I. Peddie, Veronica L. Gilrane, Eden B. King, Alexis L. & Gray, Not *So Subtle: A Meta-Analytic Investigation of the Correlates of Subtle and Overt Discrimination*, 42 Journal of Management 1588, 1592-1593 and 1599-1605 (2016).

42 *See, e.g*. Chloe FitzGerald & Samia Hurst, *Implicit Bias in Healthcare Professionals: A Systematic Review*, 18 BMC Med Ethics (2017); William J. Hall, Mimi V. Chapman, Kent M. Lee Yesenia M. Merino, Tainayah W. Thomas, B. Keith Payne, Eugenia Eng, Steven H. Day, and Tamera Coyne-Beasley (2015*). Implicit Racial/Ethnic Bias Among Health Care Professionals and Its Influence on Health Care Outcomes: A Systematic Review*, 105 American Journal of Public Health e60 (2015), https://ajph.aphapublications.org/doi/abs/10.2105/AJPH.2015.302903.

43 Lorie Fridell & Hyeyoung Lim, *Assessing the Racial Aspect of police Force Using the Implicit- and Counter-Bias Perspectives*, 44 Journal of Criminal Justice 36, 43-44 (2016).

44 *See*, Justin Nix, Bradley A. Campbell, Edward H. Byers & Geoffrey P. Alpert, *A Bird's Eye View of Civilians Killed by Police in 2015: Further Evidence of Shooter Bias*, 16 Criminology & Public Policy 309, 328-329 (2017).

45 *See,* Joshua Correll, Bernadette Park, Charles M. Judd & Bernd Wittenbrink, *The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals*, 83 J. Personality & Soc. Psychol 1314 (2002).

46 *Id*. Correll et al at 1318-1319.

47 Kristin A. Lane, Jerry Kang, and Mahzarin R. Banaji, *Implicit Social Cognition and Law*, 3 Annual Review of Law and Social Science 427, 428 (2007) (citations omitted).

48 Timothy D. Wilson, Samuel Lindsey & Tonya Y. Schooler, *A Model of Dual Attitudes*, 107 Psychological Review, 101, 120-121 (2000).

49 *See*, Madeline E. Heilman, Aaron S. Wallen, Daneilla Fuchs & Melinda Tamkins, *Penalties for Success: Reactions to Women Who Succeed at Male Gender-type Tasks*, 89 Journal of Applied Psychology 416 (2004).

50 *See*, Kristen Monroe, Saba Ozyurt, Ted Wrigley & Amy Alexander, *Gender Equality in Academia: Bad News from the Trenches, and Some Possible Solutions* 6 Perspectives on Politics 215 (2008).

men in terms of career progress and pay,[51] helping to explain the leaky pipeline, the glass ceiling, and pay inequities that occur in many professions.

Both private and public organizations have responded to this bias by introducing bias literacy training that brings these biases to conscious awareness so they can be addressed.[52] Examples found at universities include Harvard's Project Implicit,[53] the Gender Bias Learning Project,[54] Center for Worklife Law,[55] and the University of Michigan's STRIDE (Strategies and Tactics for Recruiting to Improve Diversity and Excellence).[56] Workshops often apply practices associated with adult learning and participants are taught evidence-based methods to reduce the likelihood of implicit bias.[57] Indications are that, although this training is often met with resistance, it can be effective at reducing implicit bias.[58]

Next, we consider implicit bias with respect to student evaluations.

III. Student Evaluations and Bias:

Student evaluations can contain overt bias, such as explicit statements by students that a person with a certain characteristic (e.g., gender, disability, age) should not be teaching a certain topic. However, this kind of bias is relatively rare today. More commonly now, bias is not so explicit, but arises implicitly. As the research literature indicates, even those who are explicitly supportive of equity and sure they are unbiased can demonstrate implicit bias. The types of implicit bias that could exist in student evaluations include gender, race, national origin, religion, sexual orientation, age and other dimensions that could create potential legal liability under the applicable statutes. For this paper we are not looking at the Constitutional issues under the Equal Protection Clause or the Due Process Clause, because that liberty or property interest only attaches to the right to tenure, and not to the other employment related decisions that are made using student evaluations, such as promotion, hiring decisions and merit pay increases.[59]

---

51 *See*, Alice H. Eagly & Linda L, Carli, Through the labyrinth: The Truth About How Women Become Leaders, 1-13 (2007).

52 *See*, Molly Carnes, Patricia G. Devine, Carol Isaac, Linda Baier Manwell, Cecelia E. Ford, Angela Byars-Winston, Eve Fine & Jennifer Sheridan, *Promoting Institutional Change Through Bias Literacy*, 5 Journal of Diversity in Higher Education 63 (2012).

53 https://implicit.harvard.edu/implicit/

54 https://genderbiasbingo.com/

55 Worklife Law, University of California, Hastings College of the Law, https://worklifelaw.org/

56 https://advance.umich.edu/stride/

57 *e.g.*, *see*, Sonia M. Goltz & Patty J. Sotirin, From *Academics to Change Agents in a Gender Equity Initiative*, 11 Organizational Management Journal 194, 196 (2014); Carnes, et al *supra* note 52 at 66; C. Isaac, Linda Baier Manwell, Patricia G. Devine, Cecilia Ford, A. Byars-Winston, E. Fine, Jennifer Sheridan & Molly Carnes, *Difficult Dialogues: Faculty Responses to a Gender Bias Literacy Training Program*, 21 The Qualitative Report 1243, 1257-1258 (2016).

58 *See*, Sabine Girod, Magali Fassiotto, Daisy Grewal, Manwai Candy Ku, Natarajan Sriram, Brian A. Nosek & Hannah Valentine, *Reducing Implicit Gender Leadership Bias in Academic Medicine with an Educational Intervention*, 91 Academic Medicine1143 (2016).

59 For a discussion of the Due Process issues see, Roger W. Reinsch; Susan M. Des Rosiers;

> (P)eople who seek to challenge governmental action under the due process clause must first demonstrate to the court they have a constitutionally protected liberty or property interest. If they do, and only if they do, does the court then take the next step and determine what process is due them." Therefore, not all college and university faculty members may be constitutionally protected, but for some faculty members this protected liberty or property interest does exist.[60]

A. Student Evaluations as Prompts for Bias

As discussed, the research demonstrates that the human mind functions along two very different tracks, one that generates automatic, instinctive reactions and another that produces more reflective, deliberative decisions.[61] The format of SETs, which tend to use short questions with a Likert scale, often taps into instinctive reactions instead of encouraging reflection. Additionally, many students fill these forms out in a hurry, such as at the end of class. This means that the open-ended questions that do exist and encourage reflection, which are often posted at the end of the survey, generally go unanswered or receive short responses. Therefore, the method most universities now use allows for, and even encourages, immutable characteristics such as gender, race, national origin, and age to color the results, as has been noted: "Implicit measures predict behavior to a greater extent if people do not have an opportunity to interrupt automatic processes because the behavior occurs spontaneously, or they are otherwise distracted or cognitively busy with other activities."[62]

Student evaluations also tend to ask a lot of opinion questions, creating another opening in which bias is likely to creep in. For example, here are some typical questions:

- The instructor stimulated my interest in the subject.
- The instructor demonstrated in-depth knowledge of the subject.
- The instructor appeared enthusiastic and interested.
- The instructor communicated course ideas in a clear and understandable manner.
- The instructor made it possible for me to increase my knowledge, skills, and understanding of the subject.
- My overall rating of the instruction in this course is __.

Amy B. Hietapelto, *Evidentiary and Constitutional Due Process Constraints on the Uses by Colleges and Universities of Student Evaluations*, 32 J.C. & U.L. 75 (2005).

60 Id. at 88 (citations omitted).

61 Steven A. Sloman, *Two Systems of Reasoning*, at 379 (Thomas Gilovich D. Griffin, & D. Kahneman, Eds., HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT 379-396) (2002).

62 Brian A Nosek, Carlee Beth Hawkins and Rebecca S. Frazier, *Implicit Social Cognition: From Measures to Mechanisms*, 15 Trends Cogn. Sci. 152, 156 (2011).

All these questions have the potential for implicit bias to affect the answers. The students give their opinions since these are items that ask for judgements of performance to be made without directing raters to their observation of actual behaviors. For example, "The instructor demonstrated in-depth knowledge of the subject" is strictly asking for an opinion from the student, when the student has no "in-depth knowledge of the subject" but is asked to decide whether the instructor has in-depth knowledge. All types of implicit biases may affect this answer – gender, race, age, accent, etc. The accent issue is most problematic in "the instructor communicated course ideas in a clear and understandable manner." Though most accents are perfectly understandable, they may trigger implicit bias. Therefore, this question invites the biases of students who do not want to learn to deal with the various accents they will encounter their university career.

Without anchoring these judgements in actual behaviors, expected behaviors based on stereotypes are likely to be elicited even without the rater being aware of this: "implicit and explicit social cognition exist as separate mental spheres with communication channels that are present but don't always work… Implicit and explicit measures appear to tap separate constructs that operate differently: They both predict behavior (which one predicts better appears to depend on the person and situation)."[63]

The literature on performance appraisal clearly backs up our assessment that many items in teaching evaluations are formed in a way that encourages or elicits, rather than discourages, the application of stereotypes to evaluating performance. Research indicates that performance appraisal items that are focused on behaviors or behavioral objectives tend to be more valid and less biased than measures that are more general and couched in the form of traits.[64] Asking raters to engage in a recall of behaviors has been found to reduce the impact of stereotypes on performance ratings.[65] This is because raters' focus is moved from their preconceptions to actual behaviors that were observed. Notably, organizations are more likely to be able to defend themselves in court cases when the performance appraisal instrument is behavioral in focus and has been documented to be reliable and valid.[66] The format and content of most student rating instruments, however, suggests this would be difficult to do if the use of SETs were to be challenged legally. SET items rarely ask about behaviors and they rarely ask students to recall behaviors. Also, as will be discussed more in the next paragraph, their validity is modest at best and their reliability can be questioned as well.

The research literature on student evaluations has been building substantially recently. What used to be the focus of just a few studies that generated mixed results

---

63 Kristin A. Lane, Jerry Kang, and Mahzarin R. Banaji, *Implicit Social Cognition and Law,* 3 Annual Review of Law and Social Science 427, 431-432, (2007) (parentheses in original).

64 R. Stuart Murray, *Managerial Perception of Two Appraisal Systems*, 3 California Management Review 92, 92-93(1980).

65 Cara C. Bauer and Boris B. Baltes, *Reducing the Effects of Gender Stereotypes on Performance Evaluations*, 47 Sex Roles: A Journal of Research 465, 473 (2002).

66 Gary P. Latham, Joan Almost, Sara Mann and Celia Moore, *New Developments in Performance Management*, 34 Organizational Dynamics 77, 78 (2005).

has burgeoned and effects are now clearer. Many recent studies characterized by strong research methodology have presented evidence that these biases should not be dismissed, as will be discussed in the next section. We also know much more about the characteristics of student evaluations generally. Recent advances in research methodology, such as through the examination of multilevel effects, has allowed for the separation of rating variance due to the dimensions of teachers, courses, and students.[67] These studies indicate that a large proportion of the variance in student evaluations of teaching—from 11 to 21% - is due to aspects of the students themselves rather than to aspects of teaching such as the course or instructor. Furthermore, about 25-30% of the variance results from an interaction of student and teacher characteristics. Characteristics of courses are also a strong source of variance (about 15%), meaning that, when rating teaching, students are also significantly influenced by aspects of the course the teacher cannot control.[68] These findings, in addition to the specific biases that will be discussed next, lead one to seriously question both the validity and reliability of the student ratings. In other words, the research indicates that SETs don't measure what they are intended to measure or are used for--evaluating teacher performance--because student and course characteristics play a large role, accounting for as much as 66% of the variance in ratings of instruction. This is important in that one of the key aspects courts tend to look at in performance appraisal cases is whether there is rater agreement on ratings (i.e., reliability).[69] Further, reliability is a precondition for validity, which as discussed earlier, is a factor that can affect whether defendants win court cases.[70]

B. Gender and Race Bias Effects

Gender bias has been found in student evaluations. In a recent study on gender bias in student evaluations, the researchers, Kristina Mitchell and Jonathan Martin, said, "The data are clear: a man received higher evaluations in identical courses, even for questions unrelated to the individual instructor's ability, demeanor, or attitude…Students appear to evaluate women poorly simply because they are women."[71] Other studies have had similar results and conclusions, including one by Boring-- "Female professors receive lower SET scores, despite evidence that female professors are as efficient instructors as their male colleagues"[72]--and another by Anderson and Miller--"Student expectations of the instructor, including expectations based on gender role beliefs, play a significant role in student

---

67 *Generally*, *see*, Daniela Feistauer & Tobias Richter, *How Reliable are Students' Evaluations of Teaching Quality? A Variance Components Approach*, 42 Assessment & Evaluation in Higher Education, 1263 (2017).

68 *See generally,* Feistauer & Richter *Id*. at 1273.

69 Jon M. Werner & Mark C. Bolino, *Explaining U.S. Courts of Appeals Decisions Involving Performance Appraisal: Accuracy, Fairness, and Validation*, 50 Personnel Psychology 1, 17 (1997)

70 See, Supra note 67, at 1264.

71 Kristina M. W. Mitchell & Jonathan Martin, *Gender Bias in Student Evaluations*, 51 Political Science & Politics 648, 651 (2018) (Italics in original).

72 Anne Boring, *Gender Biases in Student Evaluations of Teaching*, 145 Journal of Public Economics 27, 27 (2016).

evaluations"[73]There is also research demonstrating that the race of a professor is a factor in student evaluation results.[74] For that reason a female minority faculty member is likely to experience double the bias in SETs.[75] Professors of color have published poignant accounts of harshly negative student evaluations. The few empirical studies examining instructor race and student ratings confirm that minority faculty receive significantly lower evaluations than their White colleagues. The contradictory nature of the student comments on evaluations of minority faculty, the high levels of expressed hostility, and the occasional direct

---

73 Kristi Andersen & Elizabeth D. Miller, *Gender and Student Evaluations of Teaching*, 30 PS: Pol. Sci. & Pol. 216 (1997), arguing that "student expectations of the instructor, including expectations based on gender role beliefs, play a significant role in student evaluations" at 217; *Also see*, Susan A. Basow, *Student Evaluations of College Professors: When Gender Matters*, 87 J. Educ. Psychol. 656, (1995), finding that professor gender interacts with the gender of the student , the discipline of the course, and the questions being asked on the evaluation form at 664; Christine Haight Farley, *Confronting Expectations: Women in the Legal Academy*, 8 Yale J.L. & Feminism 333 (1996), law students exhibit gender-based stereotypes in their evaluations of female law professors; Christine M. Bachen, Moira M. McLoughlin & Sara S. Garcia, *Assessing the Role of Gender in College Students' Evaluations of Faculty*, 48 Comm. Educ. 193 (1999), a qualitative analysis to shoe how students' gender schema influences students' assessments of faculty; Marilyn S. Chamberlin & Joann S. Hickey, *Student Evaluations of Faculty Performance: The Role of Gender Expectations in Differential Evaluations*, 25 Educ. Res. Q. 3 (2001), gender can be an important factor as to how students evaluate professors; Joey Sprague & Kelley Massoni, *Student Evaluations and Gendered Expectations: What We Can't Count Can Hurt Us*, 53 Sex Roles 779 (2005), finding gender bias in an analysis of students' descriptions of instructors; Kristin A. Lane, Jerry Kang, & Mahzarin R. Banaji, *Implicit Social Cognition and Law*, 3 Annual Review of Law and Social Science 427, 439, (2007); *also see*, Linda Hamilton Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 Stan. L. Rev. 1161 (1995), One of the first law journal articles on implicit bias and employment discrimination; John W. Curtis, *Persistent Inequity: Gender and Academic Employment* (2011) at https://www.aaup.org/NR/rdonlyres/08E023AB-E6D8-4DBD-99A0-24E5EB73A760/0/persistent_inequity.pdf; Lisa L. Martin, *Gender, Teaching Evaluations, and Professional Success in Political Science*, 49 Political Science & Politics 313 (2016); JoAnn Miller & Marilyn Chamberlin, *Women Are Teachers, Men Are Professors: A Study of Student Perceptions*, 28 Teach. Soc. 283 (2000); Kristin J. Anderson & Gabriel Smith, *Students' Preconceptions of Professors: Benefits and Barriers According to Ethnicity and Gender*, 27 Hisp. J. Behav. Sci. 184 (2005), revealing in an experimental study Latina professors being more affected than male or female Anglo professors by the interactive effects of gender and ethnicity in students' ratings of professors' warmth and capability.

74 *See, e.g.* David A. Dilts, Hedayeh Samavati, Mashalah Rahnama Moghadam & Lawrence J. Haber, *Student Evaluation of Instruction: Objective Evidence and Decision Making*, 2 J. Indiv. Emp. Rts. 73 (1993, finding from student self-reports, race of instructor to be significantly correlated with student ratings; Jai Ghorpade & J. R. Lackritz, Student Evaluations: Equal Opportunity Concerns, 7 Thought & Action 61 (1991), reporting highly significant differences in student ratings favoring white over minority faculty; Katherine Grace Hendrix, *Student Perceptions of the Influence of Race on Professor Credibility*, 28 J. Black Stud. 738 (1998), suggesting that students employ different criteria to assess, and are more likely to question, the credibility and competence of Black professors than their white counterparts; Theresa A. Huston, *Race and Gender Bias in Higher Education: Could Faculty Course Evaluations Impede Further Progress toward Parity?*, 4 Seattle J. Soc. Just. 591 (2006), pointing to bias in student ratings against faculty of color; Jeannette M. Ludwig & John A. Meacham, *Teaching Controversial Courses: Student Evaluations of Instructors and Content*, 21 Educ. Res. Q. 27 (1997), demonstrating through an experimental study how race and gender interact with course content in students' expectations of professors; Deborah J. Merritt, *Bias, the Brain, and Student Evaluations of Teaching*, 82 St. John's L. Rev. 235 (2008), arguing that the conventional practices of collecting student ratings generates bias stemming from social stereotypes; Pamela J. Smith, *Teaching the Retrenchment Generation: When Sapphire Meets Socrates at the Intersection of Race, Gender, and Authority*, 6 Wm. & Mary J. Women & L. 53 (1999), discussing racial stereotypes in student ratings of teaching.

75 *See*, Andersen, etal, *supra* note 73.

references to gender or race raise troubling questions about the role of bias in these assessments.[76]

C. Other Types of Biases

In addition to race and gender, student evaluations are associated with other types of biases that fall within a category protected from discrimination. These biases include age, disability, and sexual orientation. Additionally, there are implicit bias effects that seem unrelated to gender, race, and other protected categories but that disproportionately affect certain groups. For example, studies demonstrate that attractiveness of a faculty member is a factor in SETs, leading to attractive faculty being rated nearly a full point more on a 5-point scale.[77] Even though "attractiveness" per se is not a protected category, this could easily create bias against older faculty member and disabled faculty members since they are generally viewed as being not as attractive as younger physically fit adults. Also, one study found that being attractive or not affects ratings of men more than it does ratings of women.[78] However, main effects of gender also still exist: The attractiveness study still showed that attractive women received lower ratings than attractive men.[79] This also happens in studies on age. In one study, students rated a "young" male professor higher than they rated a "young" female professor in a laboratory study that used the exact same lecture but varied the description of the professor in terms of age and gender.[80] In addition, sexual orientation, while not protected under federal law as a protected class, is protected under various city and state laws.[81] Also, recent court cases have interpreted Title VII as applying

---

76 Deborah J. Merritt, Bias, *The Brain, and Student Evaluations of Teaching*, 82 St. John's L. Rev. 234, 235-236 (2012) (Citations omitted).

77 *See, e.g.*, Daniel S. Hamermesh & Amy M. Parker, *Beauty in the Classroom: Instructors' Pulchritude and Putative Pedagogical Productivity*, 24 Econ. Educ. Rev. 369 (2005); James Felton, John Mitchell & Michael Stinson, *Web-Based Student Evaluations of Professors: The Relations between Perceived Quality, Easiness, and Sexiness*, 29 Assess. & Eval. Higher Educ. 91 (2004); James Felton, Peter T. Koper, John B. Mitchell & Michael Stinson, *Attractiveness, Easiness and Other Issues: Student Evaluations of Professors on RateMyProfessors.Com*, 33 Assess. & Eval. Higher Educ. 45 (2008); Todd C. Riniolo, Katherine C Johnson, Tracy R Sherman & Julie A Misso, *Hot or Not: Do Professors Perceived as Physically Attractive Receive Higher Student Evaluations?*, 133 J. Gen. Psychol. 19 (2006); Robert A. Lawson & E. Frank Stephenson, *Easiness, Attractiveness, and Faculty Evaluations: Evidence from RateMyProfessors. Com*, 33 Atlantic Econ. J. 485 (2005); Jennifer Bonds-Raacke & John D. Raacke, *The Relationship between Physical Attractiveness of Professors and Students' Ratings of Professor Quality*, 1 J. Psych., Psychol. & Mental Health 1 (2007); Scott Freng & David Webber, *Turning up the Heat on Online Evaluations: Does "Hotness" Matter?*, 36 Teaching Psychol. 189 (2009); Kathleen M. Silva, Francisco J. Silva, Megan A. Quinn, Jill N. Draper, Kimberly R. Cover & Alison A. Munoff, *Rate My Professor: Online Evaluations of Psychology Instructors*, 35 Teaching Psychol. 71 (2008).

78 *See generally*, Hamermesh & Parker *Id*.

79 Hamermesh, *et al Id*. See discussion at 375.

80 Julianne Arbuckle, & Benne D. Williams, *Students' Perceptions of Expressiveness: Age and Gender Effects on Teacher Evaluations*, 49 Sex Roles, 507 (2003); *also see,* William C. Levin, *Age Stereotyping College Student Evaluations*, 10 Research on Aging 134 (1988), in a study using a hypothetical male professor of three different ages (25, 53, 73), college students tended to rate the oldest professor most negatively.

81 Is Sexual Orientation a Protected Class: Everything You Need to Know, Upcounsel, https://www.upcounsel.com/is-sexual-orientation-a-protected-class.

to sexual orientation.[82] Therefore, it is important to note that research also shows that sexual orientation may have an impact on student evaluations.[83]

A related area of study that also focuses on implicit bias is the use of customer feedback by employers to make employment decisions. Even though it is debatable, students often view themselves as customers and others have also argued they should be viewed as customers.[84] Whether universities view students as customers or not, research on customer feedback is related in the sense that an employer is using third party information to make employment decisions. In a recent study of customer ratings, the authors stated:

> We set out to determine if and how customer satisfaction ratings are influenced by racial and gender biases. Across three studies we found evidence that customer satisfaction ratings are susceptible to systematic and predictable racial and gender biases. Customers tended to provide lower ratings for women and nonwhite employees and for organizations that employ such employees, than for men and white employees and their employing organizations…. Our main theoretical contributions are to show that bias appears in customer satisfaction ratings, that the bias is included in ratings of the person and the context and that it can include implicit biases. These contributions are important because they help highlight the ways and reasons that biases might appear in (any) organizational contexts.[85]

Another author stated that "Moreover, customer feedback is highly susceptible to being distorted by social group-based stereotypes and bias."[86]

Thus, the SET studies and the customer rating studies have obtained similar results and, combined, provide robust evidence that there is implicit bias in student evaluations of teaching. The next section will look at some of the discussions of implications of implicit bias for the law.

IV. Discussion of Implicit Bias and the Law

As the preceding sections indicate, research evidence is accumulating that people operate with implicit bias and that this bias shows up in different ways. Based on this type of research, Anthony G. Greenwald and Linda Hamilton

---

82 e.g. Zarda v. Altitude Express, Inc., 883 F. 3d 100 (2018) (The Second Circuit Court of Appeals ruled that Title VII protects employees from discrimination based on sexual orientation.).

83 *See*, Danny Osborne, Anne Duran, Paul G. Davies, William Wagner III, & Beth Rienzi, *Does Sexual Orientation Matter? An Experimental Assessment of Student Evaluations of a Gay Professor*, pp. 49-63 (2007) (Elizabeth M. Vargios, Ed. Educational Psychology Research Focus).

84 *See generally*, Melodi Guilbault, *Students as Customers in Higher Education: Reframing the Debate*, 46 J. of Markt. for Higher Educ. 132 (2016).

85 David R. Hekman, Karl Aquino, Bradley P. Owens, Terence R. Mitchell, Pauline Schilpzand & Keith Leavitt, *An Examination of Whether and how Racial and Gender Biases Influence Customer Satisfaction Ratings*, 53 The Acad. of Mgt. J. 238 (2010).

86 Lu-in Wang, *When the Customer is King? Employment Discrimination as Customer Service*, 23 Va. J. Soc. Pol'y & L. 249, 281 (2016)

Kriegert introduced the concept of implicit bias into the legal arena, suggesting that it has substantial bearing on discrimination law, particularly to the extent it is predictive of behavior, especially behavior diverging from avowed beliefs.[87] They noted that "evidence that implicit attitudes produce discriminatory behavior is already substantial and will continue to accumulate. The dominant interpretation of this evidence is that implicit attitudinal biases are especially important in influencing nondeliberate or spontaneous discriminatory behaviors."[88] Similarly, other authors noted: "Most important, implicit bias--like many of the heuristics and biases emphasized elsewhere--tends to have an automatic character, in a way that bears importantly on its relationship to legal prohibitions."[89]

Specifically, legal scholars note that implicit bias differs from the usual legal inquiries because many legal inquiries rely on determining the underlying intent behind a behavior or practice, whereas: "(t)he science of implicit cognition suggests that actors do not always have conscious, intentional control over the processes of social perception, impression formation and judgment that motivate their actions."[90] Therefore, the issue of intent is taken out of the picture for claims of implicit bias. This is why Greenwald and Kriegert said, "Indeed, ... implicit social cognition has the potential to influence the understanding of intent in bodies of law. For instance, constitutional and statutory law governing civil rights and the equal treatment of individuals is clearly subject to revision because implicit social cognition destabilizes conventional understandings of disparate treatment, disparate impact, hostile environments, and color or gender consciousness."[91]

Along these lines, discussions of implicit bias and the law sometimes invoke the notion of second-generation discrimination, a term introduced by Susan Sturm,[92] meaning the discrimination common today is not of the overt type typical of the first-generation discrimination cases that courts have been set up for. For example, Reinsch, Goltz and Tuoriniemi argued that second generation discrimination is not made up of the discrete intentional acts typical of first-generation discrimination that courts are more comfortable with handling, but instead is due to unconscious bias triggered by the target individual's membership in a certain group.[93] This

---

87 Anthony G. Greenwald Linda Hamilton Kriegert, *Implicit Bias: Scientific Foundations*, 94 California Law Rev.945 (2006)

88 Greenwald, etal *Id* at 961 (citations omitted); also see, Christine Jolls & Cass R. Sunstein, *The Law of Implicit Bias*, 94 Calif. L. Rev. 969, "A growing body of evidence, summarized by Anthony Greenwald and Linda Hamilton Krieger, suggests that the real world is probably full of such cases of 'implicit,' or unconscious, bias. This is likely to be true not only with respect to race, but also with respect to many other traits." at970-971 (2006) (citations omitted).

89 Jolls et al *Id* at 973.

90 Greenwald *Id* at 946; *also see*, Kristin A. Lane, Jerry Kang and Mahzarin R. Banaji, *Implicit Social Cognition and Law*, 3 Annu. Rev. Law Soc. Sci. 427 (2007).

91 Lane, etal *Id* at 439.

92 *See*, Susan Sturm, *Second Generation Employment Discrimination: A Structural Approach*, 101 Colum. L. Rev. 458, 460 (2001).

93 *See generally*, Sonia Goltz, Roger Reinsch, Joel Tuoriniemi, *University Women's Experiences in Bringing Second Generation Sex Discrimination Claims: Further Support for Adoption of a Structural Approach*, 18 Tex. J. Women & L. 145 (2009).

means that justifications for employment actions that appear on the surface to be legitimate and nondiscriminatory in reality can be justifiably questioned. In other words, an individual, group, or organization may have had the best of intentions and not ever have shown any overt discrimination, but still have been affected by implicit bias.

In addition to the above professional journal articles recognizing implicit bias, there are cases that recognize the existence of implicit bias. For example, in Adarand Constructors, Inc. v. Pena[94], Justice Ginsburg said, "Bias both conscious and unconscious, reflecting traditional and unexamined habits of thought, keeps up barriers that must come down if equal opportunity and nondiscrimination are ever genuinely to become this country's law and practice."[95] Justice Ginsburg reaffirmed that opinion in her dissent in Gratz v. Bollinger[96] by using that exact phrase again.[97] In her concurring opinion in Grutter v. Bollinger[98] she said, "It is well documented that conscious and unconscious race bias, even rank discrimination based on race, remain alive in our land, impeding realization of our highest values and ideals."[99] Justice O'Connor, in her dissent in Georgia v. McCollum[100] said, "(i) t is by now clear that conscious and unconscious racism can affect the way white jurors perceive minority defendants and the facts presented at their trials, perhaps determining the verdict of guilt or innocence."[101]

In a more recent Supreme Court decision,

> Justice Blackmun noted, discrimination has survived into our times, and is "not less real or pernicious" for "[p]erhaps... tak[ing] a form more subtle than before." Mitchell, 443 U.S. at 558-59, 99 S. Ct. 2993. The sense of a shift away from the more explicit prejudice underlying the traditional definition of discrimination has spurred the recent explosion of studies into implicit bias — that phenomenon involving the brain's use of mental associations so deeply ingrained as to operate without awareness, intention, or control. In their natural operation, implicit biases allow individuals to efficiently categorize their experiences, and these categories allow people to easily understand and interact with their world. Implicit biases can be positive or negative; it is the negative biases, however, that give rise to problems that we struggle to combat in the law and, more broadly, in our society…

---

94 515 U.S. 200, 115 S. Ct. 209,7132 L.Ed.2d 158 (1995).

95 515 U.S. at 274 (citations omitted).

96 539 U.S. 244, 123 S. Ct. 241, 1156 L.Ed.2d 257 (2003).

97 539 U.S. at 300-301.

98 539 U.S. 306, 123 S.Ct. 2325, 156 L.Ed.2d 304 (2003).

99 539 U.S. at 345.

100 505 U.S. 42, 112 S.Ct. 2348, 120 L.Ed.2d 33 (1992).

101 505 U.S. at 68; *also see*, Batson v. Kentucky, 476 U.S. 79, 106 S.Ct. 171, 290 L.Ed.2d 69 "A prosecutor's own conscious or unconscious racism may lead him easily to the conclusion that a prospective black juror is 'sullen,' or 'distant, a characterization that would not have come to his mind if a white juror had acted identically. A judge's own conscious or unconscious racism may lead him to accept such an explanation as well supported." U.S 476 at 106.

> Research has revealed the profusion of implicit attitudes that people hold towards a wide range of characteristics, chief among them the more salient and immutable traits like race and gender.[102]

With both professional journals and courts, including the Supreme Court, recognizing implicit bias, it is clear that this concept is integral to a changing legal landscape. Next, we consider the specific employment decisions that are at risk of being discriminatory because of implicit bias in SETs.

V. Specific Legal Issues Raised

As we have shown there is a risk of implicit bias being present in the answers given by student to SETs. This potential for bias raises several legal issues regarding employment discrimination. Discrimination could begin by being denied a position as a faculty member since a faculty member's student evaluations from the prior institution are often considered in the hiring process.[103] After a faculty member is hired, discrimination could occur in tenure decisions, promotion decisions and merit pay increases because SETs usually play a role in making those decisions. Poor student evaluations could be the deciding factor in whether to tenure a faculty member, which means that that faculty member is now out of a job. Merit pay decisions, although not resulting in a loss of a position, are affected more frequently by implicit bias since they generally occur yearly. Faculty in groups affected more adversely by implicit bias are likely to have lower evaluations and therefore, lower pay increases.

Thus, the result of using potentially biased SETs in hiring decisions, promotion decisions and tenure decisions could be that minority, female, and other faculty who are victims of implicit bias will not be hired, retained and/or promoted and will receive fewer rewards such as recognition and merit pay. This will result in a majority of Caucasian males who are hired, retained and/or promoted to a higher rank. This could help explain, for example, the decreasing proportions of women in academia at increased ranks that has existed for many years despite large proportions of women receiving graduate degrees as well as many efforts, such as by the National Science Foundation's ADVANCE grant program, to rectify this problem.[104] This could also help explain the pay gap between men and women

---

102 US Supreme Court Recognizes Role of Unconscious Bias in Disparate Treatment, https://www.psychologicalscience.org/news/releases/us-supreme-court-recognizes-role-of-unconscious-bias-in-disparate-treatment.html, *see*, Texas Dept. of Housing and Community Affairs v. Inclusive Communities Project, Inc., , 135 S.Ct. 2507, 192 L.Ed.2d 514 (2015).

103 *See, e.g.* Patrick Cambpell, *Student Evaluations Crucial for Hiring and Training Faculty Members,* "Each semester BYU-Hawaii students are invited to complete online class evaluations and leave comments for each of their professors, and BYUH Vice President of Academics John Bell said the school values and uses students' opinions for the hiring, retention, and classroom performance of faculty members." https://kealakai.byuh.edu/content/student-evaluations-crucial-hiring-and-training-faculty-members; *also see*, Candidate Evaluation Worksheet, d. Evidence of excellence in teaching (e.g. awards, accolades, evaluations, reviewers comments), https://advance.uncc.edu/sites/advance.uncc.edu/files/media/rubric%205_0.pdf.

104 *E.g., see*, VIRGINIA VALIAN, WHY SO SLOW? THE ADVANCEMENT OF WOMEN (1998); Abigail Stewart and Virginia Valian, An Inclusive Academy: Achieving Diversity and Excellence (2018).

that has existed in pretty much the same form since the 1970s: academic women make on average 80% of what academic men do across all disciplines, potentially resulting in over $1 million discrepancy across the lifetime of a career.[105] Over a period of several years implicit bias is likely to lead to a significant pay difference among faculty member for no other reason than some are repeated victims of implicit bias.

All these employment related decisions would violate Title VII of the Civil Rights Act of 1964.[106] Title VII of the Civil Rights Act of 1964 is a federal law that prohibits employers from discriminating against employees based on sex, race, color, national origin, and religion. It generally applies to employers with 15 or more employees, including federal, state, and local governments. Title VII forbids discrimination in any aspect of employment, including hiring and firing, compensation, promotion, recruitment, use of company facilities, fringe benefits, pay, retirement plans, and disability leave and other terms and conditions of employment.[107] In addition, there is also the Age Discrimination in Employment Act[108] which prohibits discrimination in employment against anyone over the age of 40 years old. Even though there is no specific federal legislation that prohibits discrimination on the basis of sexual orientation, the EEOC has interpreted Title VII as preventing discrimination based on gender identity or sexual orientation.[109] There are also various state laws that prevent discrimination based on sexual orientation.[110] There are two additional pieces of legislation that could apply. The Civil Rights Restoration Act of 1987 , also covers all educational institutions receiving federal funds and prevents discrimination on the basis of race, color, religion, sex, national origin, are and handicap. Finally, there is the Equal Pay Act of 1963 which prohibits pay discrepancies based on gender for substantially equal work. All of these federal laws would be relevant to situations in which SETs are used to make employment related decisions due to the types of implicit bias likely in these evaluations.

---

105 *See generally*, Kristine De Welde and Andi Stepnick, Eds., Disrupting the Culture of Silence: Confronting Gender Inequality and Making Change in Higher Education, 2015.

106 Civil Rights Act of 1964 § 7, 42 U.S.C. § 2000e et seq (1964). Title VII of the Civil Rights Act of 1964 is a federal law that prohibits employers from discriminating against employees on the basis of sex, race, color, national origin, and religion. It generally applies to employers with 15 or more employees, including federal, state, and local governments.

107 UNLAWFUL EMPLOYMENT PRACTICES, SEC. 2000e-2. [Section 703]

(a) Employer practices

It shall be an unlawful employment practice for an employer -

(1) to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, religion, sex, or national origin;

108 Age Discrimination in Employment Act of 1967 (ADEA), 29 U.S.C. § 621 (1967).

109 Sex discrimination involves treating someone (an applicant or employee) unfavorably because of that person's sex…Discrimination against an individual because of gender identity, including transgender status, or because of sexual orientation is discrimination because of sex in violation of Title VII, https://www.eeoc.gov/laws/types/sex.cfm.

110 For a list of the various protections by states as of 6/16/2019 see https://www.wisconsin.edu/lgbtq-resources/employment-non-discrimination-laws/.

Specifically, the use of student evaluations which contain implicit bias would create a claim for disparate impact because although it seems to be a facially neutral policy, as we have shown, SETs contain implicit bias. Since this is a case of unintentional discrimination, their use would be analyzed under the disparate impact framework. Even though the Civil Rights Act of 1964 did not directly address any employment policies that create a disparate impact, in Griggs v. Duke Power Co., the Supreme Court said, "The Act proscribes not only overt discrimination, but also practices that are fair in form, but discriminatory in operation."[111] This principle was codified in the Civil Rights Act of 1991,[112] which says the Act is violated when the employer engages in "a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin."[113] Thus, the use of potentially biased student evaluations to make employment related decisions would clearly fall within "a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex or national origin."

Griggs went on to say that an employment practice that does discriminate may be used if "requirements fulfill a genuine business need."[114] The Civil Rights Act of 1991 codifies this part of Griggs.[115] SETs do not need to be used to fulfill a "genuine business need". Since there are many other ways to evaluate teaching, some of which do not contain implicitly biased information, SETs are not necessary for a genuine business need. Other materials that can be used to evaluate teaching are often included in what has been called a "teaching portfolio," meaning files such as syllabi, exams and assignments, and a statement of teaching philosophy. Also, some universities use peer evaluations in which colleagues visit the classroom in addition to collecting these other materials.

VI. Recent Developments in the Use of Student Evaluations

The research with regard to bias in student evaluation has created some relatively new developments in regard to using SETs to evaluate professors. A handful of institutions have chosen not to use SETs at all to make employment decisions or to use SETs very minimally when evaluating teaching. Others have been studying the matter and generating recommendations within reports.

There are three relatively recent decisions which resulted in the stopping of the use of SETs in employment decisions altogether. These two decisions were applied to very specific institutions but have broader implications. One was an arbitrator's decision in a case involving Ryerson University in Toronto, Canada.

> An arbitration case between Ryerson University in Toronto and its faculty association that had stretched on for 15 years finally concluded with a

---

111 Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971).

112 Civil Rights Act of 1991.

113 42 U.S. Code § 2000e–2 (k)(1)(A)(i).

114 Griggs at 432;

115 "and the respondent fails to demonstrate that the challenged practice is job related for the position in question and consistent with business necessity;" 42 U.S. Code § 2000e–2 (k)(1)(A)(i).

> ruling that course surveys can no longer be "used to measure teaching effectiveness for promotion or tenure"...Arbitrator William Kaplan said that "insofar as assessing teaching effectiveness is concerned – especially in the context of tenure and promotion – SETs [student evaluations of teaching] are imperfect at best and downright biased and unreliable at worst".[116]

Granted, this is a Canadian decision, however, Philip Stark, associate dean of the Division of Mathematical and Physical Sciences at the University of California, Berkeley, who was an expert witness in the Ryerson case, said:

> (t)he impact could be much broader… Professor Stark added that he hoped that other unions representing academics in Canada, the US and elsewhere would "negotiate to reduce or eliminate reliance on student evaluations" and that universities of their own accord would "move towards more sensible means of evaluating teaching"…."I think that the time is right for class-action lawsuits on behalf of women and under-represented minorities against universities that continue to rely on student evaluations as primary input for employment decisions [and that this] will induce universities to do the right thing,"[117]

Thus, Stark was calling for unions, universities, and the courts all to take action to stop the use of SETs in making employment decisions and he was calling for this to occur internationally. Also, in 2017, the University of Southern California instituted significant changes in the use of student evaluations. This change was similar to the decision in Canada because the decision was that SETs will no longer be used in tenure and promotion decisions by the University of Southern California; however, it occurred without a union action and arbitration decision, demonstrating what Stark was calling for—voluntary action. An October 18, 2017 memo from the Vice Provost for Academic and Faculty Affairs encouraged SETs to be used to give context and provide feedback about student learning, but, "not as a primary measure of teaching effectiveness during faculty review processes given their vulnerability to implicit bias and lack of validity as a teaching measure."[118] The recommendations in this memo were then implemented by the University of Southern California:

> In a dramatic shift in faculty assessment, University of Southern California Provost Michael Quick announced that student evaluation of teaching (SETs) will no longer be an element of tenure and promotion review at the institution, Inside Higher Ed reported. Multiple studies suggest that student evaluation inherently favors white men over women and minority faculty members…. USC said it will continue to use student assessment to help professors improve their instructional design, and to shape their teaching reflection statements that will remain a part of tenure review protocols…

---

116 Ellie Bothwell, *'Tide turning' against using student evaluations to rate staff*, (July 26, 2018), https://www.timeshighereducation.com/news/tide-turning-against-using-student-evaluations-rate-staff.

117 Id.

118 https://academicsenate.usc.edu/files/2016/09/Revising-the-Student-Course-Evaluation-at-USC.pdf

> Student assessments will also be redesigned to gauge student engagement and personal responsibility taken within a course. According to Inside Higher Ed, students now will be asked about the number of hours they dedicated to course study, engagement with the professor outside of class time, and their approaches to learning course material in independent study.[119]

The University of Oregon has also made significant changes, including stopping the use of student numerical ratings in reviews and other decisions and using a more holistic approach to evaluate teaching.[120] The Oregon policy specifically states, "As of Fall 2018, faculty personnel committees, heads, and administrators will stop using numerical ratings from student course evaluations in tenure and promotion reviews, merit reviews and other personnel matters."[121]

Other universities have not taken significant action, but are studying the issue. This sometimes involves relying on recommendations from specific groups within the university that serve as task forces. For example, the University of Minnesota Women's Faculty Cabinet created a task force to look at student ratings of teaching (SRTs) to consider how they are being used across the university. In the Spring of 2019, the task force issued a report. The task force said SRT scores are being used to assess teaching performance, which impacts a variety of employment situations such as compensation and tenure. The report stated that:

> the WFC has spent the last few years investigating and compiling the strong, rigorous, and increasing evidence that SRTs are prone to bias and may have an adverse impact on women faculty, as well as faculty from other underrepresented and historically marginalized backgrounds." (Therefore), "The Cabinet has created a proposal that advocates the assembly of a diverse, university-wide and gender-balanced advisory task force to propose solutions to the SRTs are currently used, and to make suggestions on how the University can implement a more holistic evaluation process to achieve teaching excellence.[122]

Similarly, the University of Massachusetts Amherst created a faculty working group in the fall of 2017 to look at student evaluations. The working group was created to study a more robust approach to evaluate teaching and come up with recommendations. Part of the reason for this working group was that "research findings about discriminatory response biases and the sacrifice of quality for higher ratings show a complementarity of these limitations that may amplify when underrepresented faculty try to engage in novel teaching practices. These limitations in student ratings suggest that they should, at a minimum, be part of

---

119 Jarrett Carter, *USC nixes student evaluations as part of tenure review*, May 24, 2018, https://www.educationdive.com/news/usc-nixes-student-evaluations-as-part-of-tenure-review/524163.

120 Revising UO's Teaching Evaluations, https://provost.uoregon.edu/revising-uos-teaching-evaluations.

121 Colleen Flaherty, *Teaching Eval Shake-Up*, (May 22, 2018), https://www.insidehighered.com/news/2018/05/22/most-institutions-say-they-value-teaching-how-they-assess-it-tells-different-story.

122 https://faculty.umn.edu/sponsored-organizations/wfc/news.

a set of multiple measures, as is the practice when evaluating faculty research."[123] The working group came up with a proposal that was more holistic, and we will include some of those specific alternatives to SETs in the recommendations section. The University of Pittsburgh is also looking at this issue. "Provost Ann Cudd told members of Faculty Assembly on Oct. 30 that she was looking into how student evaluations are used, especially as to how they relate to the University's promotion and tenure process…This comes as the Educational Policies Committee decided in an Oct. 15 meeting to examine whether student evaluations of professors are an accurate, trustworthy measurement of teaching effectiveness. Research has found that such evaluations may hold inherent biases."[124]

This list of universities reconsidering their use of SETs is not exhaustive but is provided to demonstrate that there is broad recognition among university faculty and administrators that SETs contain bias and are problematic when the numbers are used to make employment decisions. These recent events are significant and may forecast the future in terms of the use of student evaluations. In fact, Ann Owen says "Relying on biased instruments to evaluate faculty members is institutional discrimination. Indeed, it is simply a matter of time before a class-action lawsuit is filed against an institution for knowingly using biased instruments in evaluating its faculty."[125] However, these changes might take quite some time and until most or all universities stop using SETs, we have the following recommendations as to how using them and avoid or mitigate the effects of the implicit bias they contain.

VII. Recommendations

Given their flaws, our basic recommendation is to stop using summative SETs for any employment decisions. The reason for this recommendation is that it is quite difficult to reduce the implicit bias in SETs and virtually impossible to eliminate it. For that reason, the risk of a lawsuit always exists. That does not mean that SETs would no longer exist, rather they would be used differently. For example, "SETs remain important at USC. Faculty members are expected to explain how they used student feedback to improve instruction in their teaching reflection statements, which continue to be part of the tenure promotion process… But evaluation data will no longer be used in those personnel decisions."[126]

However, admittedly, entirely dropping the use of SETs in employment decisions will be difficult to do; faculty members' job descriptions include teaching,

---

123 https://www.umass.edu/oapa/sites/default/files/pdf/program_assessment/teaching_evaluation_working_group_recommendations_finalfebruary_2018.doc.pdf.

124 http://www.utimes.pitt.edu/news/are-student-evaluations.

125 Ann Owen, *The Next Lawsuits to Hit Higher Education*, Inside Higher Ed, (June 24, 2019), https://www.insidehighered.com/views/2019/06/24/relying-often-biased-student-evaluations-assess-faculty-could-lead-lawsuits-opinion?utm_source=Inside+Higher+Ed&utm_campaign=975eb5ed6e-DNU_2019_COPY_01&utm_medium=email&utm_term=0_1fcbc04421-975eb5ed6e-199462181&mc_cid=975eb5ed6e&mc_eid=1d1653e71a.

126 Colleen Flaherty, *Teaching Eval Shake-up*, Inside Higher Ed (May 22, 2018), https://www.insidehighered.com/news/2018/05/22/most-institutions-say-they-value-teaching-how-they-assess-it-tells-different-story.

research and service, so each of those areas should be evaluated for retention, promotion, tenure, and merit pay decisions. Research is fairly easy to evaluate because there is objective evidence of the amount of research the faculty member has produced in the form of quantity of publications. Evidence of quality can be found in journal ranks, impact factors, and citation rates. The service component is also relatively straightforward--the evidence is based on the number of committees and other types of service the faculty member has participated in. That leaves the issue of evaluating teaching which is more difficult but more critical since teaching is the primary responsibility of faculty in the majority of institutions. The challenge is to evaluate teaching objectively, fairly and without bias; therefore, due to the potential for bias, the impact of SETs on employment decisions needs to be mitigated as much as possible. As discussed previously, there are a variety of methods to mitigate the impact of the implicit bias. For example, using multiple methods would create a more holistic approach for the evaluation of teaching.

In the holistic approach, SETs would continue to be used, but their impact on the employment decision would be significantly reduced. Such an approach could include such things as a peer-review model. The peer reviewers could be faculty members in the same school as the person being reviewed or they could be faculty members at other universities in the same discipline as the person being reviewed. In our experience, however, peer reviews have their own set of problems. One of the key problems encountered personally by the authors is that, in a small department, the peer reviewers may not understand the subject area. For example, one author, who teaches organizational behavior, was asked to do a peer review of an economist. The other key problem is that every faculty member knows that his/her "peers" will also review them, therefore, there review is likely to suffer from positive leniency bias. In other words, peer reviews won't truly reflect performance because they often tend to be inflated. In addition, this type of peer review may not eliminate bias, since although faculty members may know more about the dimensions of knowledgeability and effective teaching than students, they are also likely to have implicit bias in regard to gender, race, national origin, etc. To mitigate these problems, any peer reviews conducted should be behavior-based rather than trait-based and raters should be trained to avoid both implicit bias and common performance appraisal biases like leniency. An alternative to the use of internal peer reviews is to use outside peer reviewers who also teach the same course(s) at similar institutions by sending course materials including tapes of the professor teaching the course. Of course, there is a tradeoff—outside reviewers may have more expertise, but at the same time, the institution may have less control over whether they receive bias training prior to reviewing. Also, this method is likely to be time-consuming. Therefore, we suggest this be done for promotion and/or tenure decisions, since these are already time consuming and occur less often, but suggest not relying on it or not doing it as frequently when conducting merit increases.

The peer reviews in either case would not be based on sending the summative student evaluations to the peers. Instead, the reviewer would be provided with teaching materials and contextual information such as a teaching philosophy statement and told the level of each class taught, the size of each class taught, whether it is an elective or a requirement, and so forth. For example, the faculty member being reviewed would provide a syllabus of the class, the teaching

materials for the class and all the evaluation instruments used – tests, quizzes, projects, papers, etc. Then the reviewer might be told that a legal environment class is a required course for virtually every student who is majoring in some area of business and that this is typically a freshman or sophomore level class, made up of a large number of students, none of whom have an interest in that class. If the SETs are included, the instructor should be allowed to provide a written narrative about how they have, or will, respond to problematic areas. The narrative could include explanations of their teaching philosophy, why they designed the course the way they did and what they are trying to accomplish. In other words, it would be an opportunity to provide additional insight into the design and delivery of the course. The purpose of all this information would be to provide a "picture" of the class and its students to the peer reviewer so that the reviewer has some context to use for the evaluation. The point of the evaluation is to have multiple sources of data, so that the evaluator can be as objective as possible. Essentially "more information is better," a philosophy underlying the popular 360-degree feedback method.[127] Additionally, procedural justice research indicates performance appraisals are viewed as being more fair when ratees can provide input such as about important contextual factors affecting performance.[128] Perceptions about procedural fairness are associated with whether an individual is likely to see legal remedies or not.

Other recommendations for a more holistic assessment of teaching can be found in the report from the University of Massachusetts Amherst, including the following principles to guide teaching evaluation, which are paraphrased here for purposes of simplification and space:

- Evaluation should include multiple dimensions of teaching to capture the teaching endeavor in its totality, including aspects that take place outside of the classroom.

- Evaluation should include multiple sources and types of data, including faculty self-report, peer input, and student voices.

- Evaluation should involve the triangulation of measures including an acknowledgement of the ways in which these measures provide reinforcing and/or conflicting perspectives.

- Both formative and summative uses of the data should be used to maximize the impact on teaching effectiveness and a longitudinal view of teaching improvement should be taken.

- There must be a balance between uniformity across departments and customization to different disciplines.[129]

---

127 *See, generally*, Ronald A. Berk, Using the 360° Multisource Feedback Model to Evaluate Teaching and Professionalism, 31 Medical Teacher 1073 (2009).

128 *See generally*, Jerald Greenberg, *Determinants of Perceived Fairness of Performance Evaluations*, 71 Journal of Applied Psychology 340 (1986).

129 https://www.umass.edu/oapa/sites/default/files/pdf/program_assessment/teaching_evaluation_working_group_recommendations_finalfebruary_2018.doc.pdf.

Another article emphasized looking at the various learning objectives, activities, and materials and how effective they each were at generating student learning but also included considering information from the standardized evaluation form such as comments on strengths and weaknesses of the course.[130]

Essentially, what we and others are recommending is a holistic approach. As Michelle Falkoff says, what is needed are "clearer institutional policies, more mentoring of new instructors, and multiple sources of assessment. Likewise, the University of Michigan's Center for Research on Learning and Teaching emphasizes the importance of using more than one method – evaluating how faculty members deliver instructions, how they plan their courses, how they assess their students—and gathering feedback from students, colleagues, and supervisors."[131] As stated, the ideal would be to stop using SETs for any employment decisions, but if an institution decides that it must still use them, the holistic approach would at least mitigate the impact of the potential bias.

Anther method to mitigate this bias could be to use data analytics to identify student evaluations that are especially egregious. These then would be culled before calculations of mean ratings. Factors that could be flagged using either Big Data methods or other statistical processes might include whether a student tends to evaluate male professors better than female professors across time or, within an evaluation, whether a student's answers to overall dimensions do not correlate with their average ratings for more specific behaviors or traits. *Airbnb*, for example, protects hosts from inconsistent evaluations by providing reviewers with alerts when their more specific and more global ratings are inconsistent with each other and asks them if they want to change any of their ratings. Another possible mitigation strategy is to extend bias literacy training to students prior to their rating instruction, much like organizations and universities have done for faculty and staff who are involved in employment decisions. Discussions of this possibility have occurred at one of the authors' universities recently. The problem is that the student body frequently changes, so this training would require quite a bit of additional time, effort, and other resources. However, it might be important to student education more generally; therefore, some institutions might decide that it is an important investment of resources.

Ultimately, it will still be up to a court to decide how much bias is too much bias. So, if the holistic approach or other mechanisms mitigate the bias, but do not eliminate it, is that still too much bias? There are no cases that have tested this issue yet, but we contend that it is a step in the right direction. As Falkoff said "if academic institutions do not take steps to assess teaching more holistically, they run the risk of losing talented faculty members for reasons that are not only

---

130 W. Lee Hansen, *Rethinking the Student Course Evaluation: How a Customized Approach Can Improve Teaching and Learning*, https://www.aacu.org/publications-research/periodicals/rethinking-student-course-evaluation.

131 Michelle Falkoff, *Why We Must Stop Relying on Student Ratings of Teaching*, The Chronicle of Higher Education (APRIL 25, 2018), https://www.chronicle.com/article/Why-We-Must-Stop-Relying-on/243213?cid=wb&utm_source=wb&utm_medium=en&elqTrackId=173fa4ce32fa4625a8ac1fed3c6e23d1&elq=3a6dc5692327489c8a44e6bf1bf72344&elqaid=18849&elqat=1&elqCampaignId=8491.

inappropriate but may well be illegal."[132] Essentially, mitigation is a step toward both the retention of faculty who are performing effectively although SETs don't indicate that and toward avoiding possible litigation. In case litigation happens, efforts at mitigation will demonstrate that the university took substantive steps to avoid discrimination.

132 Falkoff, Id.