

# Evaluating and Retaining New STEM Faculty: How does diversity skew our “science of assessment”?

**We all like to think that we are objective scholars who judge people based entirely on their experience and achievements, but copious research shows that every one of us brings a lifetime of experience and cultural history that shapes the review process.**

Women in Science&Engineering Leadership Institute  
University of Wisconsin–Madison

- I. Evaluations: Setting the Bar
    - I. How is “the bar” gendered/racialized?
    - II. Are our “measures” scientifically suspect?
  - II. Research on gender/racial biases
  - III. Inequality Regimes & Organizational Bias
    - I. Measuring “a gentleman and a scholar”
    - II. Forms of Bias – Teaching Evaluations
  - IV. We Know a Little Moore (Research Findings)
  - V. What we can do with what we don’t yet know
    - I. Midterm evaluation–Professional consultation
    - II. Empowering Classroom Observations
    - III. Awareness and Action....
- 

# I. Setting the “bar”

- ❖ Seldin (1980) Successful Faculty Evaluation Programs  
“Informal means of evaluating faculty have always existed.”
- ❖ But recent trends: Formalize and bureaucratize evaluation methods

Current evaluation practices reflect Weber's rational bureaucratic model. National survey on use of quantitative student ratings: **88.1 percent now report they "always use" systematic numeric student ratings of teaching in evaluating teachers, up from 80.3 percent in 1988 and *only 54.8 percent in 1978.*** (Laube et al. 2007)

- ❖ Weber listed key elements of the rational–legal bureaucracy:
  1. Equal treatment
  2. Reliance on expertise, skills, and experience
  3. No extraorganizational prerogatives of position... (for personal ends).
  4. Specific standards of work and output w/ extensive records
  6. Establish/enforce rules serving organizational goals
  7. Rules and regulations mutually binding

# New faculty/New paradigms

- ▶ Guiding question: *How do women of color experience, negotiate, and persist in science and the academy? What can we all learn?*
- ▶ Changes in academic STEM culture are needed to attract and retain more diverse scientists, narrow current gaps
- ▶ The overlap between expectations for the categories of "man" and "professor" is considerably greater than for "woman" and "professor" (Martin 1984; Rakow 1991)
- ▶ Correlate higher likelihood of women and minority faculty turnover with organizational factors: deficit work environments and dominant institutional cultures
- ▶ Dr. Pardi Sabeti – Smithsonian Innovator
- ▶ Gutiérrez y Muhs, 2012. Presumed Incompetent.

# Researching the Bar for Bias

- ▶ **Women/racial minority faculty/subject to different expectations:** numbers of publications, name recognition, social networks with hiring/evaluation
- ▶ **Women's work, ideas, and findings:** may be attributed to a director or collaborator despite contrary evidence
- ▶ **Leadership ability of females/minorities:** running a research group, funding, supervision of students/staff of different gender or ethnicity may be underestimated
- ▶ **Assumptions about family responsibilities:** may negatively influence evaluation of merit, despite evidence
- ▶ **Negative assumptions about "fit":** can influence evaluation for hiring and promotion
  - ▶ University of Wisconsin–Madison

# Organizational Bias / **Inequality Regimes**

- ▶ Operate *in all organizations*
- ▶ Loosely interrelated practices, processes, actions
  - Maintain gender, racial and other inequalities

**“Even organizations that have explicit egalitarian goals develop inequality regimes over time” Acker (2006)**

- ⊗ Processes inherited, newly acquired (fluid)/ **UNEXAMINED**
  - ⊗ Linked to **steepness of hierarchy** (contrast to democratic models)
- ⊗ Gender and race segregation of jobs complex:
  - ⊗ Hierarchical across jobs at different levels, across jobs at the same level, and within jobs (Charles and Grusky 2004).
  - ⊗ Greatest in settings with **hi power, hi resources, hi competition**
- ⊗ Imbedded in bureaucratic wage setting and supervisory practices
- ⊗ Evident in interaction practices while “doing the work”

⊗ Produce *accumulating inequalities*

# In/Visibility of Inequality Regimes

- I. When evaluations focus on behaviors/attributes that resonate with race or gender stereotypes, these stereotypes *influence the standard or context used to judge a particular member of the group*
- II. Varies with the position of the beholder: “One privilege of the privileged is not to see their privilege.” Men tend not to see their gender privilege; whites tend not to see their race privilege; ruling class members tend not to see their class privilege (McIntosh 1995).
- III. Analysis of Inequality Regimes (ADVANCE)
  - Salary, Promotion and Tenure patterns, Networks of resources
  - Evaluation standards

# III. Are teaching evaluations scientifically suspect? Current Claims....

- ▶ 1980 (Seldin) – meta-review of research on *standardized SETIs* in the 1960s and 1970s
- ▶ Declared “no gender bias” in evaluations of teaching
- ▶ 1993 (Centra) – meta review
- ▶ Follow ups by Marsh, et al. 1997. Focus on “bias”
  - Student grades, classroom size, instructor age or rigor
- ▶ **No comment on SETIs for faculty of color.**

No systematic research available

- limited faculty numbers reduce statistical power
- Moore et al. 2010 – Found racialized/gendered biases in numeric student evaluations

# Threats to Reliable Evaluations

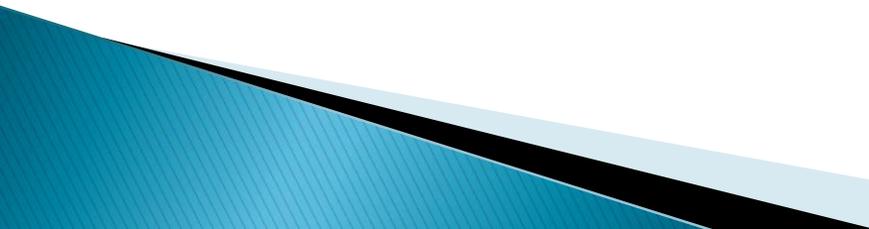
- ▶ Students **expected** and reported getting more personal time from women than from men teachers and yet were more likely to rate women teachers as not available enough (Burns–Glover and Veith 1995).
- ▶ Students' reference point for "enough availability" set at a higher order for women
- ▶ Non–random distribution of students across faculty members
- ▶ Selection bias in course enrollments\*
  - Unexamined comparisons across gender, race/ethnicity
  - Unexamined comparison across courses
    - By size, requirement, content
    - Automatic “pass” given to:
      - evaluations of Statistics courses in my major

# Threats to Valid Evaluations

- ▶ An evaluation (test) item is “valid” if it is measuring *what it is intended to measure* –
- ▶ “Dr. Fox Effect” Abrami, Leventhal, Perry (1982)
  - *Educational Seduction*
  - Instructor expressiveness has substantial effect on student *evaluations* (not achievement)
- ▶ Culturally biased threats toward privilege representatives?

# Gendered Expectations and Teaching Styles in STEM

- ▶ Expectations and masculinities in the academy
  - Competence, dominance, authority, high status
    - White male professors as “norm” (see Laube 2005)
  - Student evaluations are place-bound;
    - generated through non-inclusive practices
    - rely on standardized assumptions about teaching and learning “where white men have historically dominated definitions of excellence” (Turner 2002, 2003).
- ▶ Women academics: given lower evaluations:
  - Lower grades, cool expressive style (Sinclair 2000)
    - **More important to smile** (Kierstead et al. 1988)
  - More lecture (Statham et al. 1991)
  - Similar availability/accessibility (Bennett 1982)

- ▶ If women faculty “do gender” in a way that students disapprove, they are most at risk for angry, punitive evaluations by their students (Sprague and Massoni 2005)
  - ▶ Anecdotally: women faculty receive more comments that refer to their bodies, their physical appearance, and that objectify them sexually
  - ▶ White students may expect professors of color to behave according to preconceived constructions of racial or ethnic groups as well as view any of their criticisms of U.S. culture as "bashing" (Vargas 2002)
  - ▶ "Students tend to judge women professors first by their 'gender performance' and second by their teaching performance" (Messner 2000)
- 

## ▶ **Roots of our project**

Evaluations of diversity courses to “mainstream”  
Sociology disciplinary courses –

Racial/Ethnic Minority Faculty

Lowered by 0.9 on a 5.0 scale

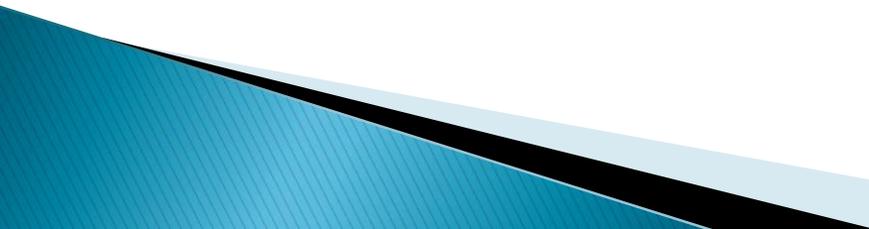
Racial/Ethnic Minority Graduate Teaching instructors

Lowered by 0.5 on a 5.0 scale

- ▶ Nationally, disproportionate assignment of instructors of color (i.e., identify as “non-white”) to diversity courses/requirements in higher education (Brayboy 2003, Perry et al. 2006).
- ▶ **Organizational blind spot** (Acker) that is difficult to unpack precisely because *the programs themselves are explicitly expected to be multicultural/global and inclusive.*

# Research Findings – Five year mixed – methods study. Drs. Acosta, Perry, Gladney

- ▶ We analyzed teaching evaluations for all required diversity classes and non-diversity courses taught at UNL: consecutive two-year time period
- ▶ Compared to “mainstream” courses / all faculty, graduate instructors
- ▶ Qualitative interviews and semi-structured surveys with faculty/GTAs
- ▶ Qualitative written student evaluation responses collected in classrooms
  - Across a wide variety of evaluation forms
- ▶ Thematically analyzed transcribed open-ended responses by students

- ▶ **Student Evaluations of courses /instructors**
  - ▶ ++“Like a jolt to the brain”
  - ▶ ++Should be required course for every student *and faculty member* at UNL
  
  - ▶ Always had a smile on
  - ▶ I didn’t sign up for this *liberal class*
  
  - ▶ Take out the nose ring
  - ▶ “bitch” “witch” and “femi-nazi”
  - ▶ Is she telling the truth about the Klan?
  - ▶ Are you some kind of black racist or something?
  - ▶ He is a *racist and I want my money back*
- 

# Unfavorable evaluations are lengthier

- ▶ “Get rid of it – diversity classes make people care less. After this class, I could care less about the plight of black people. I think if they want equality, work hard and quit complaining about the system”
- ▶ “...it was very pro-black. Overall a complete waste of time. I would not recommend anyone take this course. It made me sic ...First-like this class to teach about women in politics – as it turned out it was mostly about black women in politics”
- ▶ **Some students argued for eliminating the curriculum altogether, targeting specific faculty members**

## ▶ **Challenges to Authority:**

- ▶ For a lot of them (non-minority students), when they enter into your classroom, you're the first black person that they've ever encountered that has any sort of authority. And for some of them that's very difficult.
  - ▶ My students would be more likely to listen to what the person is saying [. . .]. They are more willing to listen; they are more receptive to white teachers. Even in the [diversity-education classroom], they are more receptive to a white person than to me.
  - ▶ Part of it is that they're dismissing both your academic and your personal experience— even though it's not necessarily your personal experience that you're presenting in the class [. . .]. I remember students were just simply unwilling to look at the evidence. And I think that's because I am a black female, (and) there is a certain element that they reject things I tell them as far as race.
- 

# Emotional Work/Burden

- ▶ What I finally learned was, “I’m not taking their rage onto me.”(LF)
- ▶ (AAF)I’m going into this place where it’s going to be me against them. It feels awful. It feels like you are before the class and you basically put on your fighting gear [. . .]. You feel like you are putting your armor on and it’s only you . . . it’s only you.
- ▶ (AAM)I don’t want to use the analogy ‘dumping’ [courses onto GTAs] . . . It’s a difficult course; people are scared
  - (AAF) Are you going to want to take on the course that is going to emotionally cost this much? And knowing that, why in the world would we give people who are green behind the ears (graduate teaching instructors) that course and then say we value that course? That’s inconsistent.

# Five words to describe your teaching experience at UNL

Thematically coded to identify modal categories

clusters by face validity strategies (e.g. “frustrated” with “exasperated” / “joy,” “delight,” “bliss”).

- ▶ **Linguistic map for women faulty of color:**
- ▶ Anger/rage/resentment, sad/depressed/discouraged, frustrated, exhausted/weary
- ▶ Frustrated/disappointed, angry, defensive/marginalized, hopeful (Men of color)

# What We Can Do about What We Don't Yet Know

- ▶ Student evaluations can tell us:
  - If the instructor is organized, prepared
  - Punctuality; at office / lab when expected
    - Norms understood? (2:30 a.m. e-mails)
  - What students observe about classroom practices
    - Instructor enthusiasm, class activities, clarity, etc.
- ▶ What can they not tell us?
  - Student attitudes, biases and beliefs
  - Their disciplinary expertise
  - Specific norms for campus/department
  - **Standardized response ranges for women/minority faculty – by discipline**

# Teaching Styles

- ▶ ADVANCE recommendations:
    - A group approach to learning often reduces competitiveness among students and reduced competitiveness may lessen or eliminate the hostility and intimidation sometimes encountered (by gender/race/ethnicity) in upper level science classes.
  - ▶ Employing various teaching styles may help foster a welcoming environment for women within the academy of engineering.
  - ▶ How are “various teaching styles” read and understood in your department?
    - Teaching observation processes
    - Student Evaluations
- 

# I. Reducing Bias: Teaching Observations

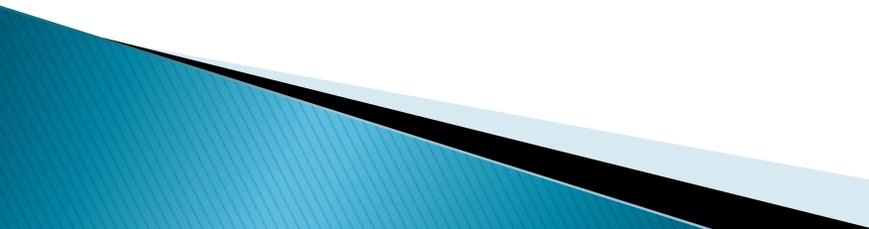
- ▶ Dreaded “Drop in” Chair model
  - Touted as “random selection”
  - Includes random interpretations (with possible biases)
- ▶ Packet on Classroom Observations
  - Training the observer
  - Gaining knowledge about your classroom
  - Reflecting that knowledge back to....
- ▶ Individual instructors might raise concerns about gender/race and evaluation in the classroom,
  - Encourage students to think critically about these practices and
  - Consider how they may be evaluating faculty differentially due to preconceived ideas about social categories (Massoni 2009).

## II. Mid-term Teaching Evaluations

- ▶ Mid-term student evaluations/feedback
  - Open ended or quantitative
  - Honest discussion and reflection
  - Yield higher summative evaluations
    - Also higher final exam scores
- ▶ Evidence that formative evaluations *can* lead to better teaching – Improvement is greatest when:
- ▶ Professional consult on interpretation
- ▶ Student evaluations include specific items (such as: professor gives preliminary overview of the lecture – in contrast to: lessons are well planned)
- ▶ Benefit: Midterm evaluations are *voluntary*
  - Formative rather than summative uses

# III. Awareness & Action (SWS Laube et al.)

- ▶ **Universities should offer/require and weigh heavily:**
  - Training in interpretation of teaching evaluations.
  - Assessment through evaluations focused on the goals and outcomes of the course
    - Reduce vulnerability to students' gendered/racialized reactions – across courses
  - Teaching portfolios, peer evaluations, and measures of teaching development activities are more reliable and valid measures of teaching effectiveness than student ratings
    - (Absent organizational inequality regimes)
- ▶ **Administrators and evaluation committees should cease using global measures: "overall, s/he is an effective teacher,"**
  - Given existing evidence of gendered and racialized perceptual bias – creates biased norms and “rubrics”

- ▶ Replace global measures with items measuring *specific practices* related to teaching effectiveness for student assessment.
  - ▶ Summaries of student evaluations should include a standard **statement acknowledging the potential impact of gender and race on class evaluations**
    - University of Mississippi
  - ▶ Institute a policy providing for the exclusion of student evaluations of teachers that are physically or sexually explicit, offensive, abusive, or unrelated to teaching
    - University of Colorado
- 

# References

- Abrami, Leventhal and Perry. 1982. Educational Seduction. *REVIEW OF EDUCATIONAL RESEARCH Fall 1982 vol. 52 no. 3 446-464*
- Gutiérrez y Muhs, Gabriella, Y. Flores Niemann, C. González and Angela Harris. 2012. *Presumed Incompetent: The Intersections of Race and Class for Women in Academia*. Utah State University Press.
- Laube, Heather, Kelley Massoni, Joey Sprague, and Abby L. Ferber. 2007. "The Impact of Gender on the Evaluation of Teaching: What We Know and What We Can Do." *NWSA Journal* 19:3 87-104.
- Marsh, Herbert and Lawrence Roche. 1997. Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52: 1187-1197
- Massoni, K. 2009 (updated). The influence of gender on student evaluations of teachers: What we can't count can hurt us." <http://www.socwomen.org/resources/genteachevalfactsheet.pdf>
- Morrisey, C and M Schmidt. 2008. Fixing the System, Not the Women. *JOURNAL OF WOMEN'S HEALTH*. 17, 1399-1408.
- Seldin, P. 1980. *Successful Faculty Evaluation Programs. A Practical Guide to Improve Faculty Performance and Promotion/Tenure Decisions*.
- Women in Science and Engineering Leadership Institute. *Research on Bias and Assumptions*. U of Wisconsin-Madison. (Packet available online)