

This article should not be reprinted for inclusion in any publication for sale without author's explicit permission. Anyone may view, reproduce or store copy of this article for personal, non-commercial use as allowed by the "Fair Use" limitations (sections 107 and 108) of the U.S. Copyright law. For any other use and for reprints, contact article's author(s) who may impose usage fee.. See also [electronic version copyright clearance](#)  
CURRENT VERSION COPYRIGHT © MMXVI AUTHOR & ACADEMIC EXCHANGE QUARTERLY

## Rhetoric, Gender and Evaluations of Teaching

Regina A. McManigell Grijalva, Oklahoma City University  
Greg R. Glau, Northern Arizona University

*McManigell Grijalva is an Associate professor of English. She researches the rhetoric of culture, ethnicity and gender; Glau is an Associate Professor of English whose research interests include higher education administration, rhetoric, composition and composition pedagogy.*

### Abstract

This study uses a rhetorical lens to reveal aspects of local context that affect how students evaluate their teachers. It also explicates how data trends can vary considerably from one institution to another, from one context to another, which suggests that the use of evaluations might be most effectively (and fairly) used for formative rather than summative purposes.

When administrators collect and use statistics, especially student evaluation numbers, it is critical to dig into the raw data and the context from which they emerge to understand what those numbers really mean. To do that, we need to examine any such data both critically and rhetorically. Rhetoric is helpful for understanding teaching and learning and students' evaluation of it because it reveals the elements within each context that affect data produced there. The subject of student evaluations has become so pervasive that The New York Times' "Learning Network" notes that high schools have now pondered the viability of using student evaluations of teaching to improve teaching (Gonchar, 2014). This increased focus on the evaluation of teaching requires a closer look at the ethics and accuracy of this method of teacher evaluation. One factor of student evaluations that has been studied by education scholars is gender and its impact on how students evaluate their teachers. The present study of teaching evaluations focuses on two large sets of student evaluations of teaching from two different institutional contexts, student populations, and time periods. This present study is only a beginning, but it is one that can reveal the rhetoricity of teaching and learning—trends distinct to each specific population. The specific findings are not necessarily transferable to other populations. The data point out that the results of student evaluations vary considerably with variance in situational factors.

Numbers produced by teaching evaluations themselves are not inherently problematic, but can be when devoid of the dynamic elements of context from whence they came. A rhetorical view would look at stable and dynamic elements of the classroom that affect how students engage in the teaching and learning taking place there and consider such complexity carefully. Generally speaking, administrators usually try to take any potential inaccuracies in data into account when making decisions based on them. This careful consideration of the possibilities for negative evaluations is an intuitive reading of teaching across several contexts. But these are minimalist ways for administrators to understand numerical data. The data in our study suggests that there may be deeper and broader reasons that affect evaluation data in negative ways, which could be why teachers often become distressed at the mention of teaching evaluations. They might be sensing that the evaluations are not accurately depicting the teaching and learning in their classrooms.

Ideas about the rhetorical situation of teaching and learning could help teachers distinguish stable classroom elements from more dynamic ones, giving teachers a sense of how to guide students toward making better learning choices among the dynamic elements, which would make students more cognizant of their learning. Administrators could also be transparent with teachers early on in teachers' courses, semesters, and careers, and help them first look at assessment formatively through a rhetorical lens. Teachers could then recognize that their communication with students shapes students' perceptions about their own and others' identities, as well as shaping actual identities. If teachers learn to read teaching and learning rhetorically, they can have more agency in assessing the classroom themselves to improve communication, which will ultimately lead to teachers who are less distressed about teaching evaluations and more confident about their own teaching.

### **Historical Trends**

It is important to reflect on the history of student evaluations of teaching when researching them or theorizing about their significance in order to understand the greater context in which their use has been developed. Thus it is important to look at what has informed or facilitated the contemporary practices of using teaching evaluations. History provides examples of stable, contextual and dynamic aspects of higher education, some slow to change and others in flux, but which affect teaching and learning over time nonetheless, including institutional missions, values and trends in teaching pedagogies, specific teaching methods and practices, and research into higher education, to name just a few. While the number of studies focusing on teaching evaluations in the past 20 years has decreased as administrators continue to shift their focus to learning outcomes (Crow, 2010), there are still many researchers interested in assessing teaching since "there is a reason to keep teaching and learning in the picture because a focus on teaching will necessarily include learning, but the inverse is not always true" (Shulman, 2004, p.16). And whatever one thinks of student comments and numerical evaluative data, that information is still part of many deliberations as to who gets a raise, who gets rehired, or whose graduate education gets funded. Thus, a rhetorical view of factors administrators consider can illuminate the tensions produced by the larger context, which come into play in summative decisions.

While there are at least five to six different categories of studies into student evaluations of teaching (Ory, 2001; Clemens Fox, 2009), and all illuminate important aspects of the teaching and learning environment, our study found how biases – in this case, what appears to be gender bias – can affect how students evaluate their teachers. Ory discussed bias studies in the larger context of studies of student evaluations of teaching, and he explained that they "attempt to identify extraneous influences on student ratings" (2001, p. 9). However, Ory was very critical of bias studies because he claimed that such influences could easily be controlled or accounted for but he never explained how in his discussion. Ory relied on the research of Basow and Silberg (1987) among others to make his claim that the "gender of students is not related to his or her ratings, although students tend to rate same-sex instructors higher" (2001, p. 6). However, the research of Basow and Silberg did not entirely support this assertion; their research revealed how complicated the picture of teaching and learning is when studies explore gender of both teachers and students and its impact on the evaluations. They looked at several different dimensions of the context of teaching and learning, and of the evaluation instrument itself and noted that on the measure of "instructor-individual student interaction," both female and male students rated female teachers less-favorably than male teachers in the study (Basow and Silberg, 1987, p. 312). Teaching and learning are complex processes as are any effective methods used to evaluate them.

There is much research to date that looks at the complexity of evaluating higher education. Some research underscores the notion that student evaluations of teaching are objective and accurate while other research supports the idea that students can be biased on their evaluations of teaching, such as the studies that have suggested that students view male and female teachers differently, and thus rate them differently. As research into the impact of gender bias progresses the body of research suggesting that gender bias does indeed affect student evaluations of teaching continues to grow (see Pritchard, 2015; Bridges, Gruys, Munsch and Pascoe, 2015; Mulhere, 2014; Boring, 2014; 2013; Young, Rush, and Shaw, 2009; Laube, Massoni, Sprague and Ferber, 2007). However, because each context differs to some degree from all others, the findings in these studies is not consistent. Researchers, Atek et.al (2015) found no gender bias factors affecting how students evaluated their teachers in their study of student evaluations of teaching. Bodhe and Jankar (2015) also found no gender bias in their study of the impact

of gender on students' evaluations of teaching. Rhetorical theory holds that each new context has its set of elements that affect the communication that transpires, so rhetoricians would expect to find such discrepancies among the results of different studies.

In this work we begin an exploration of the way that local and specific contextual elements such as gender can affect the scores teachers receive on their students' evaluations of their teaching. The notion of equal access to higher education demands that we explore such fluctuating dimensions. When administrators and teachers conceptualize the classroom and the evaluations students produce there as rhetorical with stable, relatively stable, and dynamic elements, they can begin to see what they do on many levels, at the local level and systemically in the larger context of higher education.

### **The Impact of Gender: Higher vs. Lower Evaluations Scores**

Before we turn to the two data sets in our study, it is important to look at ideas about the malleability of identity and gender. Though stable and relatively stable elements in a given context will always exist (such as classroom spaces, time and/or day of class, etc.), our study focuses on one element of the most dynamic kind, that of identity. We view the role of an actor in a communicative event (such as teaching and learning in a classroom) as subjective and fluid. An actor's identity is constructed (and reconstructed) in any communicative action in relation to other identities the actor himself/herself may assume or may confront in others within other contexts. This subjectivity of one's identity is thus open to change. Gender is a dynamic part of an actor's identity, as noted by Judith Butler (1990): "within the inherited discourse of the metaphysics of substance, gender proves to be performative—that is, constituting the identity it is purposed to be" (p. 25). Thus, our initial hypothesis was that if gender is performative, and part of one's identity, it likely has an impact on how a person evaluates another in any given context of communication, such as students evaluating their teachers. Furthermore, if males are generally more privileged in society than females and females are more often in subordinate positions than males, it is likely that female teachers are not rated as favorably as their male counterparts.

Our hypothesis, though grounded in our disciplinary focus on an English/rhetorical view of gender identity, is supported by research in other disciplines. With a sociological lens, Laube, et al. (2007) conducted a study into the impact of gender on evaluations because "gender does effect the degree of social influence individuals are perceived to have ... [and] all people tend to discount women's skills and efforts" (p. 91). Their findings suggest female teachers who do not live up to the expectations students have about teachers' performance of gender can negatively affect the students' evaluation of their female teachers.

Moreover, there are other differences among students and teachers that can skew scores from student evaluations of teachers. Philip Stark (2013), professor of statistics, notes that the numbers produced by student evaluations of teaching are very complex with regard to differentiating factors and simply cannot be simplified in the way many institutions use them: "Different courses fill different roles in students' education and degree paths, and the nature of the interaction between students and faculty in different types of courses differs. These variations are large and may be confounded with teaching evaluation scores." Finally, economics researcher, Anne Boring (2014), noted that "understanding gender biases in SET [student evaluation of teaching] scores is essential to career management in academia" (p. 3). Boring's study found that "female teachers tend to have to invest in more time-consuming dimensions of teaching [than males] if they want to improve their SET scores" (p. 30). Clearly, researchers across many disciplines have also had the sense that there are differences among students and teachers that can affect how students evaluate their teachers, gender being just one of those differences.

We wanted to address the impact of gender on how students evaluate their teachers with a rhetorical lens specific to our discipline of English rhetoric and composition. The first of our two data sets came from the main campus of one of the largest universities in the United States. As the administrator of the composition program at said institution, one of us was able to collect ten years' worth of data from the office of institutional research for the traditional ENG 101 and ENG 102 classes (first-year composition), classes that are generally required for all first-year students from academic year 1997 to academic year 2007. It is important to consider the size of this data set – roughly 900 sections, which equals close to 21,000 students. It also is important to note that the number of female composition teachers in the data

set is nearly double the number of males: 569 female “sections” and 316 male “sections” of classes are included in the data below. The data, then, reflect sections and not individual teachers. This university’s evaluations are on an inverse scale: 1 is better, while 5 is worse. The data we collected indicate that female teachers were consistently less successful on their student evaluations than their male counterparts.

Average female N=569	1.54	Average male N=316	1.46
-------------------------	------	-----------------------	------

On average, then, females received worse evaluations than their male counterparts: 1.54 for female teachers as compared to 1.46 for male teachers. While the average difference between females and males was .09, if there would be a change in the elements of the rhetorical situation, there would also be a change in the numbers that evaluate them.

Our hypothesis that “female teachers get worse evaluations than male teachers” statement was also true for ENG 101 and ENG 102:

ENG 101 avg. female	1.53	ENG 101 avg. male	1.41
ENG 102 avg. female	1.59	ENG 102 avg. male	1.49

On a percentage basis, female teaching evaluations were about 8% worse in ENG 101 and about 6% worse in ENG 102.

Many foundational discussions of assessment in English rhetoric and composition or writing studies, tend to strongly point out the importance of local context to the assessment of teaching and learning (for example, see Broad, 2003; O’Neill, Moore, and Huot, 2009; and Broad, et al., 2009, Detweiler and McBride, 2009). Teaching and learning are situational, as they happen always within the context of communication. The elements that make up the context of communication, such as audience, purpose, and context, vary from one larger context to another, such as from one institution to another, or even from one local context to another, such as from one class to another. Any shifting of the elements, and the ways that students perceive the teaching and/or learning is bound to reflect such shifts.

Moreover, identities are unstable, and are often constructed within the very act of communication itself, such as in the moment of teaching and learning in a classroom. Factors that shape perceptions of identity, such as gender, are bound to play a part in many contexts. Other factors that could impact perceptions of teaching and learning include the people, places, and things within the classroom, the exigence of the communication taking place there, and the meaning constructed within the act of communication. Ideas about the shifting contextual elements in communication make it clear that evaluations must always be read as contextual; they must look largely at the local elements of the specific teaching and learning situation for them to reveal what is happening there, and, more importantly, for them to improve the efficacy of teaching and learning. With regards to the first analysis of the data, many questions about identity, context, and even purpose surfaced.

This first data set of student evaluations of teachers who taught first-year composition brought the question up of if/how gender affected the way students related to and evaluated their teachers in the English composition classroom. The results suggested that gender biases could have a statistically significant impact on scores teachers received on their teaching evaluations – but why were female evaluations worse than males’? And—how was the department “reading” these evaluations, and what impact, if any, did the discrepancy have on how teachers were evaluated by the English Department? That is, should it take into consideration any anti-female-teacher bias, if in fact it did exist?

There are so many elements of teaching and learning that affect evaluations scores that in light of the research, and of the conversations happening among teaching assistants, we continued the investigation into what the numbers said, especially considering the high stakes of teaching evaluations in summative

decisions. And even though many researchers in education and other fields outside rhetoric and composition or writing studies put much credence into statistics and quantitative data, we wanted to join others in the rhetoric/composition/writing discipline who were urging educators and scholars to “re-capture, re-coup, and harness organic, localized assessment to nourish productive teaching and learning” (Broad, et al., 2009, p. 2). That is not to say that the numbers are not valuable, but that they are also only a small part of the total picture of what happens in the composition classroom.

An analysis of the context and how important it is to the rhetorical situation of teaching was in order. Female teachers from the first data set were rated more unfavorably than their male counterparts at this first specific campus location, so the question emerged of if/how those numbers might change given a different demographic of students and/or teachers, and a different teaching/learning context. Teaching and learning are important to view in terms of the teachers and students themselves, but they should never be viewed separately from their context. There have been many disagreements among education scholars over the years as to whether bias factors actually affect the results of student evaluations. Scholars such as Lombardo and Tocci (1979) and Bennett (1982) have researched gender bias and found the difference it made to be insignificant, while others such as Basow (1995), and Bernstein and Burke, and Centra (1995) have found bias factors such as gender to be significant. One finding common to all of these studies was that female students rated both female and male teachers similarly, but male students tended to rate female teachers less favorably than their male counterparts. Results of such studies should differ when an institution’s specific demographic data and programs differ (which is usually the case). The present study confirmed this to be so; demographics and context can and do change the numbers significantly.

The first data set seemed to suggest that female teachers get, on average, worse evaluations than their male counterparts. But that changed with an analysis of data set from an institution with different demographics and programs than those of the first data set. Here a higher number is better.

Female Teachers average = 3.99  
N=136

Male Teachers average = 3.89  
N=99

However, on average, the second data set showed that female teachers received better evaluations than their male counterparts: 3.99 for female teachers as compared to 3.89 for male teachers. While the average difference for female teachers and male teachers was .10, this is roughly a 3% difference, which is statistically significant. The results from both data sets confirmed what our rhetorical theories suggested about the importance of carefully reading the elements of each specific context of teaching and learning to better understand the numbers that student evaluations of teaching produce in the composition classroom.

The second data set of teaching evaluations confirmed what rhetorical theories about communication and emergent literature on assessment in rhetoric/composition suggest. The institutional research office at the second institution gave us access to the second data set of teachers’ evaluations scores and their genders from 2007 – 2010. The student population at the second institution is about one-third the size of the student population at the first data set institution, so the evaluation data set is much smaller. For a six-semester period (235 sections of first-year composition), on average male teachers got worse evaluations than their female counterparts—exactly the opposite of the data from the initial university campus study.

There could be many reasons for this difference, including the fact that the teachers from the first data set teach the traditional ENG 101-102 sequence of classes, while the teachers from the second data set have a one semester requirement: its four-credit hour ENG 105 class. In examining these data, though, one interesting fact stood out: the student population at the two universities, in terms of gender, was remarkably different. At both universities – as is typical throughout the country – female students outnumbered male students. (Note that for this comparison, we assume that the student population in the ENG classes mirrors the overall student population at each university.)

The percentage differences in terms of overall student population look like this:

	Female undergraduates	Male undergraduates
1st Institution	52%	48%
2nd Institution	60%	40%

At first glance, those percentages are not that far apart . . . but when we consider how many female students and how many male students there would be in a typical class, the difference is striking. In a typical class of, say, 25 students, a teacher would have:

	Female undergraduates	Male undergraduates
1st Institution	13 students	12 students
2nd Institution	15 students	10 students

When this difference is considered, at the second institution where the second set of data were derived, there are five more females evaluating a teacher than there are males. At the first institution where our initial data set were derived, there is only one more female evaluating a teacher than there are males doing so. This suggests that the marked difference in demographics, such as gender and general education requirements for first-year composition, affected the difference in the outcome of the teachers' evaluations.

The first data set was very large (1300 sections taught by approximately 200 teachers), certainly well above the statistically significant sample size that probability theory under the Central Limit Theorem (which is 30) would suggest is large enough for statistical significance (Rice, 1995), were we relying strictly on quantitative methods. As noted above, there is much value in numbers, but only when they are explained within their local and specific contexts, with their rhetorical elements made visible. The sample from the second institution, which was compared to the initial sample was also large enough to be considered statistically significant (235 sections). These findings continue to resonate with what contemporary researchers are saying about gender in the field of student evaluations of teaching, such as Pritchard (2015), Boring (2014), and Stark (2013).

### **What to Do Next?**

As noted above, some research has found that gender among other factors affects how students evaluate the learning environment. Also noted above, other research has found that bias had little or no impact on how students evaluated their teachers. A careful review of student evaluations research shows that the context from one data set to the next varies enough for the results to reflect that variation. Rhetoricians have understood the importance of specific context to communication for centuries. The differences from one teaching and learning context to another simply cannot be ignored or disregarded when the stakes are high, when the numbers the teaching evaluations produce could be the reason a teacher is denied promotion or tenure, or is not rehired; or when the numbers are the reason a graduate teaching assistant cannot continue to fund her graduate education with stipends for teaching. Developing a culture of transparency and quality in higher education should be bottom-up, unilateral, and top-down, working in a classroom between teachers and students; working across groups of teachers; and in the greater institution between/among administrators and teachers. Once institutions, programs, colleges, and/or departments are clear about their own cultural contexts of teaching and learning, they can determine what their specific situational expectations are with regard to student evaluations of teaching.

### **Conclusion**

Gender is an important factor to consider in the evaluation of teaching, but it is only one of many. Our research provides a glimpse into how disparate numbers from evaluations can be across two different contexts. It is a microcosm of the larger set of studies into student evaluations of teaching undertaken in the past century. These studies are increasingly more important as educational institutions seek to evaluate the quality of teaching. However, the more administrators look for ways to quantify the quality of something as complicated as teaching and learning, the more research must look at the complexity of what transpires in the classroom to ensure accuracy. Variance does not mean the numbers are wrong, but it does suggest that it would be unethical to use student evaluations data for a summative decision without a close look at the context and culture from which the numbers emerge. Most importantly, it tells

us that formative uses for student evaluations data could lead to greater communication among teachers and their students, and to improved growth and development of educators.

## References

- Atek, E. S., Salim, H., Halim, Z. A., Jusoh, Z., & Yusuf, M. A. (2015). Lecturer's Gender and Their Valuation of Student Evaluation of Teaching [Electronic version]. *International Education Studies*, 8(6).
- Basow, S. A. (1995). Student evaluations of college professors: When gender matters. *Journal of educational psychology*, 87(4), 656-665.
- Basow, S. A., & Silberg, N. (1987). Student Evaluations of College Professors: Are Females and Male Professors Rated Differently?. *Journal of Educational Psychology*, 79(3), 308-314.
- Bennett, S. K. (1982). Student Perceptions and Expectations for Male and Female Instructors: Evidence Relating to the Question of Gender Bias in Teaching Evaluation. *Journal of Educational Psychology*, 74(2), 170-179.
- Bernstein, B., & Burke, L. (1995). Contributions of Faculty, Student, and Course Characteristics to Student Evaluations: Report on Student Evaluations of Female and Male Faculty at Arizona State University. B. L. Bernstein & L. Burke with S. Blaisdell (Eds.). Tempe: Arizona State University Faculty Women's Association.
- Bodhe, C. D., & Jankar, D. S. (2015, January). Teaching Effectiveness: How Do Students Evaluate their Teacher? [Electronic version]. *International Journal of Healthcare and Biomedical Research*, 3(2), 155-159.
- Boring, A. (2014, April 6). Gender Bias in Student Evaluations of Teachers. In Anne Boring.
- Bridges, T., Gruys, K., Munsch, C., & Pascoe, C. J. (2015, March 24). Beyond Bossy or Brilliant: Gender Bias in Student Evaluations of Teaching. In *The Society Pages*.
- Broad, B. (2003). *What We Really Value: Beyond Rubrics in Teaching and Assessing Writing*. Logan: Utah State University Press.
- Broad, B., Adler-Kassner L., Alford B., Detweiler J., Erstrem H., Harrington S., McBride M., Stalions E., & Weden S. (2009). *Organic Writing Assessment: Dynamic Criteria Mapping in Action*. Logan: Utah State University Press.
- Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge, Chapman, & Hall.
- Centra, J. A. (2000, Jan/Feb). Is There Gender Bias in Student Evaluations of Teaching? *The Journal of Higher Education*, 71(1), 1-22.
- Crow, M. M. (2010, May). *Toward Institutional Innovation in America's Colleges and Universities*. *Trusteeship Magazine*, 2-5.
- Clemens Fox, R. (2009). *Teaching and Learning, and Their Evaluation, Made Visible: Using Rhetorical Theories to Assess Higher Education*. La Vergne: VDM Verlag.
- Detweiler, J. & McBride, M. (2009). *Designs on assessment at UNR: University of Nevada at Reno. Organic Writing Assessment: Dynamic Criteria Mapping in Action*. Logan: Utah State University Press.
- Gonchar, M. (2014, Sept. 5). *Should Students Be Able to Grade Their Teachers?* *The Learning Network of The New York Times*. Retrieved 26 December 2015.
- Laube, H., Massoni, K., Sprague, J., & Ferber, A. L. (2007). The Impact of Gender on the Evaluation of Teaching [Electronic version]. *National Women's Association Journal*, 19(3), 87-104.
- Lombardo, J., & Tocci, M. E. (1979). Attribution of positive and negative characteristics of instructors. *Perceptual and Motor Skills*, 491-494.
- O'Neill, P., Moore, C., & Huot, Brian. (2009) *A Guide to College Writing Assessment*. Logan: Utah State University Press.
- Ory, J. C. (2001). Faculty Thoughts and Concerns about Student Ratings. *Techniques and Strategies for Interpreting Student Evaluations*. K. G. Lewis. (Ed.), (pp. 3-15). New York: Jossey-Bass.
- Pritchard, S. B. (2015, June 23). *Should Female Faculty Get Bonus Points to Correct for Gender Bias?* In *Higher Ed Jobs*. Retrieved 21 December 2015.
- Rice, J. (1995). *Mathematical Statistics and Data Analysis*, 2nd ed. North Scituate: Duxbury Press.
- Shulman, L.S. (2004). Visions of the possible: Models for campus support of the scholarship of teaching and learning. *The scholarship of teaching and learning in higher education: Contributions of research universities / edited by Becker, W. E. and Andrews, M. L.* Bloomington; Indiana University Press.
- Stark, P. (2013, Oct. 14). *Do Student Evaluations Measure Teaching Effectiveness?* *The Berkeley Blog*. University of California Berkeley. Retrieved 23 December 2013.
- Young, S., Rush, L., & Shaw, D. (2009). Evaluating Gender Bias in Ratings of University Professors [Electronic version]. *International Journal for the Scholarship of Teaching and Learning*, 3(2)19, 1-14.