DOCUMENT RESUME

ED 387 848 CS 509 061

AUTHOR Boggs, Cathy
TITLE Identifying Gender Bias in Teaching Evaluations: Insights Offered by Communication Research.
PUB DATE 28 May 95
NOTE 35p.; Paper presented at the Annual Meeting of the International Communication Association (45th, Albuquerque, NM, May 25-29, 1995).
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Classroom Communication; College Faculty; *Communication Research; Higher Education; *Instructional Effectiveness; *Sex Bias; *Student Attitudes; *Student Evaluation of Teacher Performance; Teacher Student Relationship
IDENTIFIERS Gender Issues; Research Suggestions

ABSTRACT

This paper examines insights offered by communication research that may be valuable to researchers looking for evidence of gender bias affecting student ratings of their college professors' teaching effectiveness. The paper offers an overview of the teaching evaluation processes, and conceptualizations of the validity of their measures. It then discusses evidence that students' evaluations of teaching effectiveness may be subject to a number of biases, including gender bias. It then reviews research indicating that classroom communication processes are influenced both by professor and student gender. It is suggested that findings from two lines of communication research, instructional communication and gender and communication, may provide insights into the processes through which communication contributes to gender biases in student ratings. The paper concludes with specific and general suggestions for using these insights in future research. (Contains 71 references.) (Author)

Identifying Gender Bias in Teaching Evaluations: Insights Offered by Communication Research

by Cathy Boggs
University of California
1832 Ellison Hall
Santa Barbara, CA 93106-4020
(805) 893-4479
e-mail: 6500cbog@ucsbuxa.ucsb.edu

Paper presented to the Instructional and Development Division of the International Communication Association, Albuquerque, New Mexico, May 28, 1995

Abstract

This paper examines insights offered by communication research that may be valuable to researchers looking for evidence of gender bias affecting student ratings of their college professors' teaching effectiveness. The paper offers an overview of teaching evaluation processes, and conceptualizations of the validity of their measures. It then discusses evidence that student evaluations of teaching effectiveness may be subject to a number of biases, including gender bias. It then reviews research indicating that classroom communication processes are influenced both by professor and student gender. It is suggested that findings from two lines of communication research, instructional communication and gender and communication, may provide insights into the processes through which communication contributes to gender biases in student ratings. The paper concludes with specific and general suggestions for using these insights in future research.

Student evaluations of their professors' teaching serve as key factors in many universities' hiring, promotion, and retention procedures, and influence the pay and promotional opportunities

of university instructors at all levels of the academic hierarchy (Cruse, 1987; Sandler, 1991). Yet although they have been widely used for most of the century (McKeachie, 1990), the reliance of university administrators on student evaluations to estimate teaching effectiveness has always been controversial because of questions about their validity: no one is certain what they measure (Cruse, 1987; Stewart & Roach, 1993).

Critics have charged that student ratings are subject to a great many potential biases unrelated to teaching effectiveness (Braskamp, Brandenberg, & Ory, 1984; Cruse, 1987; Wigington, Tollefson, & Rodriguez, 1989). Among the loudest critics are feminist scholars, who charge that social and institutionally-entrenched sexism affects teaching evaluations, and thereby places women faculty at a significant professional disadvantage in relation to their male colleagues (Sandler, 1991; Sekaran & Kassner, 1992; Wood & Lenze, 1991). However, studies examining gender effects on teaching evaluations have found mixed results: some studies have appeared to support bias charges, whereas others have not (Basow & Silberg, 1987; Bennett, 1982; Marsh, 1984; Wigington et al., 1989).

In viewing the results of these studies from the perspective of a communication scholar, it appears that one reason for researchers' difficulty in determining the existence, or lack thereof, of gender bias in student teaching evaluations is the lack of a communication focus in such research. Although most researchers recognize that teaching is essentially a communication process (Marsh, 1984; Nussbaum, 1992b), most of the existing studies on gender bias in teaching evaluations have

been conducted by educational and social psychologists with little knowledge of communication theory or research findings. As a result, such studies have lacked recognition of the dynamics by which communication factors influence the quality of classroom interactions, as well as their effects on evaluations of those interactions by participants.

The goal of this paper, then, is to suggest ways that insights provided by the communication literature can assist researchers in identifying productive routes to investigating gender bias effects on university teaching evaluations. To accomplish this goal, the paper first offers an overview of teaching evaluations and discusses considerations related to their validity raised by their supporters and critics. Next the paper discusses potential biasing factors affecting student ratings of teacher effectiveness, and summarizes the findings of studies examining gender bias on those ratings. Research findings indicating that gender factors influence communication between college teachers and students are then summarized. Following this, the paper discusses two distinct lines of communication research, instructional communication and gender and communication, that offer insights into the ways that gender and communication factors may have biasing influences on social evaluation of teachers. The paper then offers suggestions for incorporating these insights into research efforts examining gender bias in teaching evaluations, and concludes with a discussion of the implications of these suggestions for researchers studying academic gender bias.

Uses and Nature of Teaching Evaluations

Student evaluations are a widely-used method for rating the teaching effectiveness of college professors in the U.S., Canada, and other nations (Marsh, 1984; Sallinen-Kuparinen, 1992). Quantitative teaching evaluations are collected primarily for two purposes: (1) to provide diagnostic feedback to faculty, and (2) as measures of teaching effectiveness on which to base tenure and promotion decisions (Marsh, 1984; Stewart & Roach, 1993), although they may also be collected for the purposes of (3) offering guidance to students on course selection or (4) providing outcome data for educational research
(Marsh, 1984). While college instructors, like other people, are generally resistant to being evaluated, their resistance to teaching evaluations is primarily related to their function as data for personnel decisions (Cruse, 1987; Marsh, 1984; Stewart & Roach, 1993).

One of the major criticisms of teaching evaluations is that the information they provide is atheoretical. The construct "teaching effectiveness" can be operationalized in many different ways, depending largely on the interests of the evaluators and the purposes for which the data are used (Cashin & Downey, 1992; Marsh, 1984, 1991; Stewart & Roach, 1993). Teaching evaluation forms are often poorly designed, consisting of items purported to measure teaching effectiveness that are chosen without regard for teaching or learning theories explaining their relationship to particular student learning outcomes (Abrami, d'Apollonia, & Cohen, 1990; Marsh, 1984).

Braskamp et al. (1984) describe three general styles of

teaching evaluation forms used on American college campuses. Omnibus forms are those given to all members of the faculty across departments within the university, and contain items representing major areas of instruction believed by members of that academic community to be important components of effective teaching. Omnibus forms usually contain items relating to some or all of the following topics: communication skill, rapport with students, course organization, student self-rated accomplishments, course difficulty, and grading and examinations. Goal-based forms request information on students' perceptions of their mastery of knowledge related to course goals and objectives, such as gaining factual knowledge, developing special skills, and developing appreciation for course subject matter. Cafeteria system forms are those created by using items drawn from a mix-and-match pool of items supplyinginformation for campuswide, departmental, and specific course evaluation purposes.

Teaching evaluation forms require different levels of information from student raters and focus attention on different aspects of the teaching process. Items may require students to make inferences about global characteristics of professors ("How would you rate this instructor overall?"), general characteristics (" The teacher seemed well-prepared") , and specific characteristics ("The teacher answered difficult questions clearly"). (Braskamp et al., 1984). Items may address process variables (those related to global teaching methods and specific teaching behaviors), presage

variables (characteristics of teachers and students), product variables (student academic or professional achievement and attitudes), and context variables (substantive, physical, and institutional environments) (Dunkin & Barnes, 1986; Erdle, Murray, & Rushton, 1985). Teaching evaluations used on college campuses consist primarily of items addressing process and product variables; items dealing with presage and context variables generally are not examined except in studies addressing ratings validity questions (Marsh, 1984).

While there is great variation in the form and content of teaching evaluations, studies of student ratings have tended to define effective teachers as being high in factors such as Enthusiasm/Dynamism, Organization/Clarity, Individual Rapport, Group Interaction Skills, and Grading Fairness (Cadwell & Jenkins, 1985; Leventhal, Perry, & Abrami, 1977; Marsh, 1984), all but the last factor related wholly or in part to communication skills.

Validity of Student Evaluations of Teacher Effectiveness

Validity of student evaluations has been conceptualized in two different ways: as accurate reflections of students' opinions about their instructors, or as accurate indicators of instructors' teaching effectiveness (Abrami et al., 1990). While supporters of the use of student ratings claim that they successfully fulfill both functions (Abrami et al., 1990; Marsh, 1984; McKeachie, 1990), opponents claim they are actually more useful as indicators of student consumer satisfaction than teaching effectiveness

(Abrami et al, 1990; Cruse, 1987).

Despite the lack of guiding theories behind the development of teaching evaluation forms, evaluation proponents point to several methods that can be used to determine the validity of student teaching evaluations. Construct validity can be established by showing that items related to the same factor are more closely correlated with each other and less correlated with items designed to measure other factors (Kremer, 1990; Marsh, 1984). For instance, responses on items measuring instructor dynamism and enthusiasm for the course material should be more closely related to each other than items measuring grading fairness . Another method for determining the validity of student evaluations is to compare them to ratings results on other criterion measures of teaching effectiveness, such as student test scores, changes in student behaviors, instructor self-evaluations, observational data collected by peers and/or administrators, and the findings of experimental research (Howard et al., 1985; Kremer, 1990; Marsh, 1984; McKeachie, 1990). If former students asked to rerate their professor one or more years after taking a class provide ratings similar to the original ones, the stability of those ratings may be considered an indication of their validity (Howard et al., 1985; McKeachie, 1990). Also, high reliability levels among ratings by students in the same class can be considered evidence of the validity of students' observations (Marsh, 1984).

Opponents of the use of student evaluations do not

consider proponents' approaches to testing ratings' validity as sufficient to meet their concerns, however. Without an adequate theory basis, it is not possible to know whether the factors described as being associated with effective teaching truly serve as valid measures of that construct (Cruse, 1987). For the same reason, convergent and divergent correlations among related questionnaire items, while adequate to show construct validity, cannot be considered evidence for the external validity of teacher ratings: even when item correlations indicate that students rate their teachers similarly along the same dimensions, they do not reveal why students judge their teachers as they do (Cruse, 1987). Just as student ratings of teaching effectiveness may be suspect, so are criterion measures from other sources; student achievement test scores are more highly correlated with prior student ability than with teacher behaviors (Cruse, 1987), and peer evaluations may partially reflect the rater's knowledge of the professor's past student ratings (Murray et al., 1990; Stewart & Roach, 1993). Critics also charge that various bias factors influence student evaluations, so that ratings may actually measure presage characteristics of the teacher and contextual characteristics of a course rather than actual features of the teaching processes that they are supposed to represent (Cruse, 1987).

Research into bias effects on student ratings of professors suggests that critics' charges of non-teaching influences on evaluations may be warranted, and that feminist claims of gender

Marsh also suggests that such correlations should not be considered as evidence of bias if those factors are shown to affect student learning outcomes. Yet because measures of student learning are not perfectly correlated with teaching quality, further research is needed to confirm the validity of Marsh's latter recommendation.

Research has found evidence that a number of personal and contextual factors unrelated to teaching behaviors may influence student's ratings of their professors. Among these factors are course characteristics such as class size (Cranton & Smith, 1990), academic level (Murray, Rushton, & Paunonen, 1990), subject matter (Marsh, 1984), presentation format (Murray et al., 1990), and even the timing of evaluations (Abbott, Wulff, Nyquist, Ropp, & Hess, 1990). Teacher characteristics found to influence ratings include traits such as enthusiasm (Feldman, 1986), extraversion (Erdle et al., 1985), achievement orientation (Murray et al., 1990), instructor expressiveness (the so-called "Dr. Fox effect") (Abrami, Leventhal, & Perry, 1982; Ware & Williams, 1980); instructor rank (Leventhal et al., 1977; Cranton & Smith, 1990), and instructor's perceived attitude similarity to the student (Abrami & Mizener, 1985). Student characteristics found to influence ratings include students' age (e.g., adult learners vs. traditional undergraduates, as in Comadena, 1992), use of implicit theories about correlations between seen and unseen characteristics of professors (Marsh, 1984), and use of general rather than specific judgments of

bias may not be easily dismissed. However, while claims of bias tend to win wide support from sceptics of the use of student ratings in teacher pay and promotion processes (Marsh, 1984), the conflicting results of such studies does not yet offer conclusive proof that the ratings are invalid.

Biases Affecting Teaching Evaluations

As is the case with other teaching evaluations research, studies of bias factors affecting student ratings have been criticized for the lack of theory informing them. Marsh (1984) has suggested that most studies of evaluation biases simply assume that bias exists when they find correlations of presage and context variables with process variables. Abrami et al. (1990) suggest that such correlations among teacher characteristics, course characteristics, and instructional style characteristics should only be considered as evidence of bias if they are found to have no effects on product measures of instructional effectiveness, such as student test scores, which are often not included in ratings bias studies. Conversely, they consider findings showing high correlations of potential biasing factors with product variables and low correlations with student ratings to be valid evidence of bias; such would be the case if women professors' students achieve higher test scores but they receive lower student ratings than their male peers. Marsh (1984) says that bias may be indicated by statistical correlations of non-teaching factors, such as class size, with measures on some, but not all, categories of multidimensional student ratings.

teacher characteristics leading to halo effects (Cadwell & Jenkins, 1985 ; Cook 1989; Cruse, 1987). None of these findings has been unequivocally supported across multiple studies.

Gender Bias Effects on Teaching Evaluations

Researchers investigating the effects of gender on student ratings have studied ways in which teacher sex and student sex influence evaluation outcomes. Like other researchers investigating biases in teaching evaluations, investigators of gender bias have found mixed results. Few studies of gender differences in teaching evaluations of professors show clear-cut differences in ratings due to teacher gender alone, and other studies have shown no gender differences in student ratings (Basow & Silberg, 1987). However, some evidence of potential gender bias has been found.

Some studies have found that men and women college instructors may be rated according to different criteria for teaching effectiveness. Several studies have found that student ratings of women faculty are tied to perceptions of how often they smile and how sociable they are perceived to be, whereas these factors were considered to be much less important for male instructors (Hall, Braunwald & Mroz, 1982; Kierstead, D'Agostin, & Dill, 1988). Bennett (1982) found that women teachers were judged to be more professional if they used a highly structured approach in presenting course material, whereas men's professionalism ratings were unaffected by this factor. In this same study, women's credibility in the classroom was closely tied

to the professionalism ratings, as well as to ratings of self-assurance and compellingness. Consistent with these findings, Hall et al. (1982) found that women whose self-presentation in the classroom was traditionally feminine were judged as less competent than women who did not exhibit such behaviors.

Other research indicates that male and female professors may be rated differently according to the same performance criteria. Bennett (1982) found evidence that students hold women instructors to higher standards than men with regard to how much time they are expected to spend in supportive activities: in her study, women instructors were rated as less accessible than men even when they spent more time working with students outside the classroom. Cooper, Stewart, & Gudykunst (1982) found that perceived instructor concern for the student's welfare was the strongest predictor of both women and men instructors' evaluations. However, they also found that men's ratings were closely tied their competence in delivering formal feedback (perceived fairness of grades awarded), whereas women's ratings were related instead to their ability to provide informal feedback reinforcing students' self-concepts (in the form of written evaluations of student performance).

An experiment by Basow & Distenfeld (1985) found that for female teachers, a nonexpressive teaching style led to high ratings, while the same style led to low ratings for males instructors. Wigington, Tollefson, & Rodriguez (1989) found that men teaching upper division classes received higher student

evaluation ratings than women professors teaching similar classes at the same university.

Bennett (1982) found that women teachers were rated as more effective overall than men, and that this difference was related to the greater perceived warmth and charisma of women instructors in the interpersonal aspects of their teaching (willingness to assist, encouraging expression, and ability to arouse and sustain interest). In line with these findings is research indicating that men who display traditionally male professorial behaviors are rated as scholastically rigorous and intellectually challenging, whereas women professors utilizing similar behavior are rated as rigid and controlling (Basow & Silberg, 1987; Sandler, 1991).

Student sex has also been found to be a potentially important factor in evaluations given to college professors. Basow & Silberg (1987) found a significant teacher sex by student sex interaction on ratings of four dimensions of teaching effectiveness, with male students rating female professors significantly lower than male professors on measures relating to scholarship, organization/clarity, dynamism/enthusiasm, and overall teaching ability. Kaschak (1978) found that men students consistently rated male professors as more effective, concerned, likeable, and excellent than female professors, whereas female students gave men and women professors equivalent ratings on these measures. Winocur, Schoen, & Sirowatka (1989) found that Australian female students preferred affiliative teaching styles

over instrumental styles, whereas the male students displayed no preference either way.

In contrast to the above findings is research indicating that gender bias may not be a factor in student evaluations. Bennett (1982) found that men and women received similar ratings from their students on criteria of egalitarianism in interpersonal style (e.g., warmth), personal charisma, self-assurance, and instructional approach. In addition, both Bennett (1982) and Basow & Silberg (1987) have pointed out that in many studies where gender differences were found, effects sizes were quite small and do not offer strong support for the existence of gender bias.

Part of the difficulty that researchers have had in establishing firm evidence about the existence or nonexistence of such biases may be that little of this research has been informed by communication theory or the findings of communication research (Marsh, 1984; Stewart & Roach, 1993), despite the obvious role that communication processes play in their creation and transmission. Therefore, the following sections of the paper suggest ways that communication perspectives might assist researchers in formulating studies of gender bias by providing models for analyzing relevant educational communication processes. First, the findings of studies showing some of the ways that gender has been found to influence classroom communication patterns are reviewed. Then several models are presented that may be useful in analyzing the communication-

related processes involved in creating gender bias in student teacher evaluations, followed by suggestions for their use by researchers.

Gender and Communication in the College Classroom

Research on communication patterns in schools below and at the university level offer evidence that gender has significant influences on patterns of classroom interaction. A number of studies illustrate that instructors use different communication styles with male and female college students. Studies have found that college teachers of both sexes call on female students less frequently to answer questions in class (Pearson & West, 1991), make more disparaging comments to women and discourage women from classroom participation (Hall & Sandler, 1982), and provide male students with more opportunities to interact than female students (Sadker & Sadker, 1985).

While both male and female professors engage in communication behaviors that are different in relation to women and men students, there is also evidence that men and women instructors exhibit different teaching styles. In a review of research on gendered talk in the academy, Treichler & Kramarae (1983) cite research findings that women instructors encourage more classroom participation by students than do men, and that men tend to be more direct and women less direct in offering criticism to students.

There is also evidence that teacher and student sex interact in their influences on faculty-student interactions. A

study by Brooks (1982) found that male graduate students in mixed-gender classes tended to speak more frequently and for longer turns, and to interrupt female students and their professor more frequently in classes taught by women than in classes taught by male professors. Sandler & Hall (1986) found that female professors experience frequent challenges to their authority and qualifications by male students not experienced by their male colleagues. On the other hand, female students have been shown to be more willing to speak up in a class taught by a female professor than a male one (Karp & Yoels, 1976; Pearson & West, 1991). Sandler (1991) found that women students communicate more with women professors, while Pearson & West (1991) found that men students asked more questions in classes taught by men professors than in classes taught by women.

Studies conducted by communication researchers suggest that gendered communication practices such as those described above may influence students' evaluations of their teachers both in terms of teaching and gender factors; therefore, both types of influences must be taken into account in searching for patterns of gender bias. A number of communication perspectives provide models of social evaluation processes that offer relevant insights into the dynamics of how these influences may affect student ratings. Reviewed below are insights offered two lines of communication research that are highly relevant to these issues: instructional communication and gender and communication.

Insights from Research on Instructional Communication

Whereas much educational research tends to be uninformed by findings from communication research, studies on instructional communication usually examine communication processes related to factors that education researchers have identified as important to classroom outcomes (Stewart & Roach, 1993). Instructional communication researchers have developed two models for evaluating teaching effectiveness that may help in illuminating the ways that gender considerations affect student ratings.

Norton's (1978, 1983) model of communicator style has been used by a number of instructional communication researchers as the basis for studies attempting to develop a profile of effective teaching (Nussbaum, 1992b; Sallinen-Kuparinen, 1992). According to Nussbaum (1992b), Norton (1978) defines communicator style as "the way one verbally and paraverbally interacts to signal how literal meaning should be taken, interpreted, filtered, or understood." (p. 146). The model suggests that both content (message) and relational (style) elements of teacher communication influences student performance outcomes and evaluations (Sallinen-Kuparinen, 1992). According to Nussbaum (1992b), the model defines four characteristics of teacher communicator style: it is observable (can be identified by observers), multifaceted (individuals can utilize many style subconstructs), multicollinear (behaviors within each dimension are correlated with each other), and it is

variable but sufficiently patterned (an individual has many choices of communication behavior in a given situation, but tends to choose consistently from a restricted range of behaviors over time). Teacher communicator style differs from individual communicator style in that it is constrained by requirements related to the teaching role, such as course content, level of the class, and class size (Nussbaum, 1992b).

Kearney & McCroskey (1980) extended Norton's model by integrating it with concepts of assertiveness, responsiveness, and versatility as defined in the social style construct of Buchholz, Lashbrook, and Wenberg (1975) (Sallinen-Kuparinen, 1992). Rather than conceiving of teacher communicator style as a set of objectively-observable communicative features, Kearney and McCroskey defined teacher communicator style as "the collective perceptions of others and/or self-perceptions of a teacher's relational image in the classroom," (Kearney & McCroskey, 1980, p. 533). This conceptualization of teacher style thus takes into account the relational communicative competence aspects of classroom communication that are ignored in most process-product models of teaching evaluation used by educational researchers (Sallinen-Kuparinen, 1992).

Researchers using teacher communicator style approaches have examined the effects of a variety of teacher behaviors on teacher effectiveness ratings, including teacher immediacy, dramatics, nonverbal expressiveness, disclosiveness, delivery skills, time spent in contact with students, dynamism, and liking

for students (Nussbaum, 1992a). They have found that effectiveness ratings are related to perceptions of teachers' overall communicative ability (Nussbaum, 1992b), use of immediacy behaviors (Anderson & Anderson, 1987), use of humor, self-disclosure and narratives (Nussbaum, Comedena, & Holladay, 1978). Summarizing the results of this research, Nussbaum (1992b) describes effective teachers as those who rate themselves and whose students rate them as competent communicators. Specific characteristics that correlate with these ratings include being dramatic, relaxed, friendly, and impression-leaving. Behaviors that correlate with these ratings include moving about the room, engaging in eye contact, using humor in lectures, self-disclosing and telling narratives to clarify course content.

Stewart & Roach (1993) suggest that Richmond & Gorham's (1992) Instructional Communication Model offers a framework for understanding the communicative process central to teaching and learning, and so may be useful for studying teaching evaluation validity. Components of the model are the teacher, course content, instructional strategy, the student, and evaluation/feedback, all operating within a specific learning environment/context. Within the model, teacher and student function as communicative interactants, each bringing their own experience, biases, personality, and communicative styles to a given interaction. Teachers and students provide evaluation and feedback to each other on various learning objectives, and about relational aspects of the interaction.

Stewart & Roach (1993) state that all components of the model can influence teaching evaluations, including the context/environmental features which are frequently ignored in teaching evaluations research. They suggest that by adopting a communication-based model such as Richmond and Gorham's, it is possible for researchers to make predictions about expected relationships between model components. As a result, evaluation biases resulting from the interaction of communication variables with those related to teacher and student characteristics would be made visible.

Insights from Research on Communication, Gender, and Social Evaluation

Researchers studying gender differences in social evaluations of communicators have found evidence that such differences are related to two factors involved in the communication process: (1) differences in communicative characteristics between men and women speakers, and (2) listeners' sex role stereotypes about what is gender-appropriate behavior. Different communication researchers tend to emphasize one factor over the other in their work. Reviews by Aries (1987), Hall (1987), Haslett (1993), and Smythe (1991) describe the findings of the numerous studies that have addressed both of these factors.

Researcher who examine gender differences in evaluation outcomes linked to communication characteristics base their work on evidence that men and women may use somewhat different

patterns of communication behavior. Linguists and communication researchers have identified a numerous examples of apparent verbal and nonverbal differences in men's and women's conversational styles, including their use of tag questions (Lakoff, 1973), question-asking and backchannelling (Maltz & Borker, 1982; Tannen, 1990), gaze (Mulac, Studley, Wiemann, & Bradac, 1987), and qualifiers (McMillan, Clifton, McGrath, & Gale, 1977). These findings have been used in the development of a "gender as culture" model of communication. The model posits that women and men's speech differences contribute to the development of separate male and female value systems, and that these in turn lead to differences in the social evaluation of men's and women's communication contributions.

One line of communication research that shows the ways that gender differences in communicative behaviors may be related to differences in the social evaluations received by men and women communicators is the gender-linked language effect (GLLE) studies conducted by Mulac and associates. Studies of the GLLE have found that differences in men's and women's linguistic choices lead to social evaluation outcomes in which women are evaluated as being higher than men on dimensions of socio-intellectual status and aesthetic quality, while men are rated as being more dynamic than women. These findings have been shown to be robust in a variety of conditions, including research with written transcripts (Mulac, Incantro, & James, 1985), television programs (Mulac, Bradac, & Mann, 19850, public

speeches (Mulac & Lundell, 1982), and in dyadic interactions (Mulac, Wiemann, Widenmann, & Gibson, 1988).

While research has found that the interpersonal communication styles of men and women may differ in some respects, there is also evidence that, overall, women and men's communication behavior is more similar than different (Eagly & Johnson, 1990; Smythe, 1991). Yet research has found evidence that even when male and female communicators utilize similar communication behaviors, the sex-role stereotypes of their listeners may result in different interpretations, and evaluations, of their communicative contributions. Studies by Goldberg (1968) and Paludi and Strayer (1985) found that subjects rate texts purportedly written by men as higher in quality and value, and rate the author as significantly more knowledgeable and professional, than when the identical texts are presented as having been written by women. Paludi and Strayer found that while ratings by both men and women subjects showed the same pattern of favoring male authors, men subjects rated women authors significantly lower than did women subjects.

Recommendations for Researchers

Thus far, this paper has discussed evidence that student evaluations of their college instructors' effectiveness may be subject to a number of biases, including gender bias. It has reviewed research indicating that classroom communication processes may be influenced both by professor and student gender, suggesting that communication may provide the key to

understanding the source and manifestation of gender biases in teaching evaluations. Also discussed were two lines of communication research that offer insights into the processes through which communication may contribute to gender biases in evaluation: instructional communication and gender and communication. The following section of the paper suggests ways that insights from these perspectives can be combined to provide a conceptual basis for investigating gender bias in student teaching evaluations.

Taken together, findings on teacher communication style and those on communication, gender and social evaluation suggest that teacher characteristics and features of the instructional environment may combine with gender considerations to influence the nature of the teacher-student relationship, and thus bias teaching effectiveness ratings. To find evidence of gender bias in the evaluations process, if it exists, researchers must therefore investigate student ratings in relation to the interactions among four factors: (1) women and men instructors' classroom teaching behaviors; (2) students' classroom behaviors in courses taught by men and women professors; (3) students' expectations for the performance of their male and female instructors; and (4) contextual characteristics of the learning environment.

Insights offered by the two communication perspectives provide indications not only of the factors which should be studied, but how they should be investigated. Both approaches

offer evidence that certain communication behaviors used by professors and students are more closely related to biasing effects on evaluations than others: specifically, those that have been identified as important contributors both to teaching evaluations and to gender-related social evaluations, such as dynamism. More importantly, the two communication approaches offer researchers guidance in identifying the types of linkages among the four factors and ratings outcomes that would be expected to exist if gender biases are involved. For example, differences in overall student ratings of men's and women's teaching effectiveness could be attributed to gender bias effects if they are found to be consistently associated with factors such as instructors' use of gender-appropriate teaching behaviors, students' expectations of how gender-appropriate teaching behaviors correlate with instructor knowledgability, differences in the amount and quality of students' levels of interactions with male and female professors, and differences in the characteristics of courses taught by men and women (such as class size, presentation format, and course content). Examples of the types of research questions suggested by this conception of gender bias are the following:

Do men and women teachers' use of communication behaviors differ along the dimensions predicted to be important to teaching evaluations by the teacher communicator style models?

Are there differences in the ratings that students give to men and women on measures of dynamism/enthusiasm and intellectual status, as would be predicted by the gender-linked language effect?

Are the questions asked in teaching evaluation instruments gender-biased, because they emphasize dimensions of teaching effectiveness that are differently related to stereotypes for appropriate masculine and feminine behaviors?

Do contextual factors of classes, such as presentation style, class size, and academic level, interact with sex role stereotypes to influence perceptions of appropriate communicator behaviors for men and women instructors?

Do correlations of course content, use of particular communicator style variables, and teacher effectiveness evaluations differ for men and women professors?

In what ways do student's communicative behaviors influence teachers' communicator styles?

Do stereotypes related to women's and men's communication styles lead to use of implicit theories by students, or result in ratings biased by halo effects?

Conclusions

In addition to suggesting specific questions for future research, the above discussion suggests that researchers addressing gender bias in student teaching evaluations keep several important caveats in mind when designing their studies.

First, studies addressing gender bias in evaluations need to consider the interactional aspects of the teaching process. Rather than conceptualizing teaching evaluations as outcomes directly related to teacher behavior, such studies need to take into account the transactional processes by which students and teachers mutually influence each other's classroom performance (Nussbaum, 1992a; Stewart & Roach, 1993).

Secondly, researchers should look for effects of institutional factors upon student expectations about gender-appropriate communication behavior For example, in many universities the role of college professor is still primarily filled by male professors. If women teachers who present themselves in the classroom use behaviors appropriate to the professorial role, they may appear to violate expected standards of femininity, and thus be rated lower by students than male professors displaying the same teaching style (Sandler, 1991; Sandler & Hall, 1986).

Thirdly, decisions about the presence or absence of gender bias in student ratings cannot be made on the basis of statistical evidence alone. Statistical tests such as those indicating convergent or divergent measures of the construct validity of ratings are only as good at finding bias as the items included on the instrument permit (Cruse, 1987).

Fourth, this discussion shows the need for communication and education researchers to work more closely together to develop study designs investigating gender bias that better incorporate the strengths offered by both perspectives. Not only would educational researchers benefit by using more communication concepts in their work, but communication researchers would benefit by following the lead of educational researchers in conducting more observational research inside actual classrooms. (Nussbaum, 1992a).

Finally, this discussion suggests that researchers should cease the quixotic quest to develop the perfect teaching

evaluation instrument. No one-size-fits-all measure of effective teaching has been or is likely to be found at anytime in the near future. While research on communicator style has shown that ratings of teacher effectiveness are related to the use of behaviors connoting dynamism, openness, and drama (Nussbaum, 1992b), evidence is growing that no single teaching style is effective in every situation (Sprague, 1993). Factors that correlate positively with effective teaching outcomes under some circumstances are found to correlate negatively or not at all in others: for instance, Sallinen-Kuparinen (1992) found that, in contrast to results of U.S. studies showing a significant negative correlation between authoritarianism and positive evaluations in ratings by non-college students, the two factors were not significantly related for students in her Finnish sample.

In conclusion, it appears that student teaching evaluations, while good reflections of students' opinions about their teachers, may be less valuable as objective measurements of teaching effectiveness due to their vulnerability to a wide range of hard-to-identify biases, including those related to the gender of professors and students. Yet despite their limitations, use of student evaluations in personnel processes can be expected to increase as university administrators come under increasing pressure from state legislators and parents to prove the "value" of university teaching (Hay, 1992). It is thus more important than ever to find out what they actually measure.

Braskamp, L.A., Brandenburg, D.C., & Ory, J.C. (1984). Evaluating teaching effectiveness: A practical guide. Newbury Park: Sage.

Brooks, V. (1982). Sex differences in student dominance behavior in female and male professors' classrooms. Sex Roles, 8, 683-690.

Cadwell, J., & Jenkins, J. (1985). Effects of the semantic similarity of items on student ratings of instructors. Journal of Educational Psychology, 77, 383-393.

Cashin, W.E., & Downey, R.G. (1992). Using global student rating items for summative evaluation. Journal of Educational Psychology, 84, 563-572.

Comadena, M.E., Semlak, W.D., & Escott, M.D. (1992). Communicator style and teacher effectiveness: Adult learners versus traditional undergraduate students. Communication Research Reports, 57-63.

Cook, S.S. (1989). Improving the quality of student ratings of instruction: A look at two strategies. Research in Higher Education, 30, 31-45.

Cooper, P.J., Stewart, L.P., & Gudykunst, W.B. (1982). Relationship with instructor and other variables influencing student evaluations of instruction. Communication Quarterly, 30, 308-315.

Cranton, P., & Smith, R.A. (1990). Reconsidering the unit of analysis: A model of student ratings of instruction. Journal of Educational Psychology, 82, 207-212.

Cruse, D.B. (1987). Student evaluations and the university professor: caveat professor. Higher Education, 16, 723-737.
Dunkin, M.J., & Barnes, J. (1986). Research on teaching in higher education. In M.C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 754-777). New York: Macmillan.

Eagly, A.H., & Johnson, B.T. (1990). Gender and leadership style: A meta-analysis. Psychological Bulletin, 108, 233-256.

Erdle, S., Murray, H.G., & Rushton, J.P. (1985). Personality, classroom behavior, and student ratings of college teaching effectiveness: A path analysis. Journal of Educational Psychology, 77, 394-407.

Feldman, K.A. (1986). The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: A review and synthesis. Research in Higher Education, 24, 139-213.

Goldberg, P.A. (1968). Are women prejudiced against women? Transaction, 5, 28-30.

Hall, J.A. (1987). On explaining gender differences: the case of noverbal communication. In P. Shaver & C. Hendrick (Eds.), Review of Personality and Social Psychology, Vol. 7: Sex and Gender (pp. 177-200). Newbury Park, CA: Sage.

Hall, J.A., Braunwald, K.G., & Mroz, B.J. (1982). Gender, affect, and influence in a teaching situation. Journal of Personality and Social Psychology, 43, 270-280.

Hall, R., & Sandler, B. (1982). The classroom climate: a chilly one for women? Washington, D.C.: Project on the Status and Education of Women, Association of American Colleges.

Hall, R., & Sandler, B. (1986). Out of the classroom: a chilly campus climate for women? Washington, D.C.: Project on the Status and Education of Women, Association of American Colleges.

Haslett, B. (1993). Gender, power, and communication in organizations. In B. Dervin & U. Hariharan (Eds.), Progress in communication sciences XI (pp. 159-177). Norwood, NJ: Ablex.

Hay, E.A. (1992). A national survey of assessment trends in communication departments. Communication Education, 41, 247-257.

Howard, G.S., Conway, C.G., & Maxwell, S.E. (1985). Construct validity of measures of college teaching effectiveness. Journal of Educational Psychology, 77, 187-196.

Hurt, H.T., Scott, M.D., & McCroskey, J.C. (1978). Communication in the classroom. Reading, MA: Addison-Wesley.

Karp, D., & Yoels, W.C. (1976). The college classroom: some observations on the meanings of student participation. Sociology and Social Research, 60, 421-439.

Kaschak, E. (1978). Sex bias in student evaluations of college professors. Psychology of Women Quarterly, 3, 235-243.

Kearney, P., & McCroskey, J.C. (1980). Relationships among teacher communication style: Trait and state communication apprehension and teacher effectiveness. In D. Nimmo (Ed.), Communication yearbook 4 (pp. 533-551). New Brunswick, NJ: Transaction Books.

Kierstead, D., D'Agostin, P., & Dill, H. (1988). Sex role stereotyping of college professors: bias in student rating of instructors. Journal of Educational Psychology, 80, 342-344.

Kremer, J.F. (1990). Construct validity of multiple measures in teaching, research, and service and reliability of peer ratings. Journal of Educational Psychology, 82, 213-218.

Lakoff, R. (1973). Language and woman's place. Language in Society, 2, 47-48.

Leventhal, L. Perry, R.P., & Abrami, P.C. (1977). Effects of lecturer quality and student perception of lecturer's experience on teacher ratings and student achievement. Journal of Educational Psychology, 69, 360-374.

Maltz, D.N., & Borker, R.A. (1982). A cultural approach to male-female miscommunication. In J.J. Gumperz (Ed.), Language and social identity (pp. 196-216). Cambridge: Cambridge University Press.

Marsh, H.W. (1984). Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases, and utility. Journal of Educational Psychology, 76, 707-754.

Marsh, H.W. (1991). A multidimensional perspective on students' evaluations of teaching effectiveness: Reply to Abrami and d'Apollonia (1991). Journal of Educational Psychology, 83, 416-421.

McKeachie, W.J. (1990). Research on college teaching: The historical background. Journal of Educational Psychology, 82, 189-200.

McMillan, J.R., Clifton, A.K., McGrath, D., & Gale, W.S. (1977). Women's language: Uncertainty or interpersonal sensitivity and emotionality? Sex Roles, 3, 545-559.

Mulac, A., Bradac, J.J., & Mann, S.K. (1985). Male/female language differences and attributional consequences in children's television. Human Communication Research, 11, 481-506.

Mulac, A., Incantro, C.R., and James, M.R. (1985). Comparison of the gender-linked language effect and sex role stereotypes. Journal of Personality and Social Psychology, 49, 1098-1109.

Mulac, A., & Lundell, T.L. (1982). An empirical test of the gender-linked language effect in a public speaking setting. Language and Speech, 25, 243-256.

Mulac, A., Studley, L.B., Wiemann, J.M., & Bradac, J.J. (1987). Male/female gase in same-sex and mixed-sex dyads· Gender-linked differences and mutual influence. Human Communication Research, 13, 323-343.

Mulac, A., Wiemann, J.M., Widenmann, S.J., and Gibson, T.W. (1988). Male/female language differences and effects in same-sex and mixed-sex dyads: The gender-linked language effect. Communication Monographs, 55, 315-335.

Murray, H.G., Rushton, J.P., & Paunonen, S.V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. Journal of Educational Psychology, 82, 250-261.

Norton, R. (1978). Foundation of a communicator style construct. Human Communication Research, 4, 99-112.

Norton, R. (1983). Communicator style: Theory, applications, and measures. Beverly Hills, CA: Sage.

Nussbaum, J.F. (1992a). Effective teacher behaviors. Communication Education, 41, 167-180.

Nussbaum, J.F. (1992b). Communicator style and teacher influence. In V.P. Richmond & J.C. McCroskey (Eds.), Power in the classroom: communication, control, and concern (pp. 145-158). Hillsdale, NJ: Lawrence Erlbaum.

Paludi, M.A., & Strayer, L.A. (1985). What's in an author's name? Differential evaluations of performance as a function of author's name. Sex Roles, 12, 353-361.

Pearson, J.C., & West, R. (1991). An initial investigation of the effects of gender on student questions in the classroom: developing a descriptive base. Communication Education, 40, 22-32.

Richmond, V.P., & Gorham, J. (1992). Communication, learning, and affect in instruction. Edina, MN: Burgess International Group, Inc.

Sadker, M., & Sadker, D. (1985). Sexism in the classroom. Vocational Education Journal, 60, 30-32.

Sallinen-Kuparinen, A. (1992). Teacher communicator style. Communication Education, 41, 153-166.

Sandler, B.R. (1991). Women faculty at work in the classroom, or, why it still hurts to be a woman in labor. Communication Education, 40, 6-15.

Sandler, B.R., & Hall, R.M. (1986). The campus climate revisited: Chilly for women faculty, administrators, and graduate students. Washington, D.C.: Project on the Status and Education of Women. Association of American Colleges.

Scott, M.D., & Nussbaum, J.F. (1981). Student perceptions of instructor behaviors and their relationship to student evaluation. Communication Education, 30, 44-53.

Sekaran, U., & Kassner, M. (1992). University systems for the 21st century: proactive adaptation. In U. Sekaran & F.T.L. Leong (Eds.), Womanpower: managing in times of demographic turbulence (pp. 163-191). Newbury Park: Sage.

Smythe, M.J. (1991). Gender and communication behaviors: A review of research. Progress in Communication Sciences X (pp. 173-215. Norwood, NJ: Ablex.

Sprague, J. (1993). Retrieving the research agenda for communication education: asking the pedagogical questions that are "embarassments to theory." Communication Education, 42, 106-120.

Staton-Spicer, A.Q., & Wulff, D.H. (1984). Research in communication and instruction: categorization and synthesis. Communication Education, 33, 377-391.

Stewart, R.A., & Roach, K.D. (1993). A model of instructional communication as a framework for analyzing and interpreting student ratings of instruction. Communication Quarterly, 41, 427-442.

Tannen, D. (1990). You just don't understand: women and men in conversation. New York: Ballantine.

Treichler, P.A., & Kramarae, C. (1983). Women's talk in the ivory tower. Communication Quarterly, 31, 118-132.

Ware, J.E., & Williams, R.G. (1980). A reanalysis of the Dr. Fox experiments. Instructional Evaluation, 4, 15-18.

Wigington, H., Tollefson, N, & Rodriguez, E. (1989). Students' ratings of instructors revisted: Interactions among class and instructor variables. Research in Higher Education, 30, 331-344.

Winocur, S., Schoen, L.G., & Sirowatka, A.H. (1989). Perceptions of male and female academics within a teaching context. Research in Higher Education, 30, 317-329.

Wood, J.T., & Lenze, L.F. (1991). Strategies to enhance gender sensitivity in communication education. Communication Education, 40, 16-21.