

## **Student Evaluations of College Professors: Identifying Sources of Bias\***

---

**KENNETH M. CRAMER & LOUISE R. ALEXITCH**

*University of Saskatchewan*

### **ABSTRACT**

Previous studies have found that students' evaluations of their professors' teaching ability may be affected by such factors as students' expectations and gender stereotypes. The present study examined how students' evaluation of faculty may be affected by student's gender, professor's gender, discipline, and a variety of demographic and social variables. Undergraduate students ( $N = 910$ ) evaluated one of their professors on sensitivity to students' needs, quality of teaching, course structure, and treatment of designated group members (e.g., visible minorities). Results showed that female students rated their professor higher on sensitivity to students' needs and treatment of designated groups than male students. Science students rated their professor lower on teaching quality and treatment of designated groups than either Social Science or Fine Arts/Humanities students. In addition, students' ratings correlated with how often a professor met with students outside of class, when the class was scheduled, and class size. The implications for using student evaluations to accurately assess professors' teaching ability are discussed.

---

\* This study was funded by a President's Social Science and Humanities Research Council Grant (#7-70405) awarded to both authors. The sequence of authorship was randomly determined and thus both authors deserve equal merit for this work. Special thanks is given to both R. Paola Lake and Steven Lake for data collection.

## RÉSUMÉ

Des études antérieures ont démontré que les évaluations faites par les étudiants au sujet de l'aptitude à l'enseignement de leurs professeurs peuvent être affectées par des facteurs tels que les attentes des étudiants et les stéréotypes sexuels. Cette étude examine comment l'évaluation des étudiants de leur faculté peut être affectée par leur genre, celui du professeur, la discipline, et une variété d'autres variables démographiques et sociales. Des étudiants de premier cycle ( $N = 910$ ) ont évalué un de leurs professeurs sur la sensibilité de celui-ci à leurs besoins, sur la qualité de l'enseignement, sur l'organisation du cours, et sur le comportement à l'égard des membres de certains groupes particuliers (e.g., les minorités visibles). Les résultats ont montré que les étudiantes, en comparaison avec les étudiants, ont évalué plus fortement leurs professeurs quant à la sensibilité à leurs besoins et au comportement à l'égard des membres de groupes particuliers. Les étudiants en Sciences en comparaison avec ceux des Beaux Arts et des Lettres ont évalué plus faiblement la qualité de l'enseignement et le comportement à l'égard des membres des groupes particuliers. De plus, les évaluations des étudiants étaient en corrélation avec le nombre de fois où un professeur s'est trouvé avec les étudiants en dehors du cours, la place occupée par le cours à l'horaire, et la taille du groupe. Les conséquences de l'utilisation des évaluations d'étudiants pour mesurer avec précision la capacité d'enseigner de leurs professeurs sont discutées.

---

Some review articles (e.g., Marsh & Roche, 1997; McKeachie, 1997) expounding upon the validity of student evaluations of their professors, have determined that students' ratings reflect a multidimensional construct, that variables such as expected — actual grade, class size, and instructor's rank have mixed effects on students' ratings, and that (for the most part) students' ratings are valid indicators of professors' teaching. While the present authors do not dispute these conclusions, it is conceivable that many of these extraneous variables may have interactional effects on students' evaluations (e.g., professor gender and student gender), and

that students' perceptions of the overall institutional climate may affect their ratings of professors.

According to Baird (1990) and Pascarella (1984), the interplay among people, processes, and institutions creates a climate that guides the behaviour of its constituents. In either college or university settings, the climate includes the perceptions, expectations, satisfactions, and dissatisfactions of the people who make up the campus community, such that an individual may behave uniquely or feel treated distinctly in such a climate. Early investigations into campus climate assessed the nature of student gender bias (Harvey & Hergert, 1986; Williams, 1990), and revealed a learning environment less favorable to women (Carelli, 1988; Heller, Puff, & Mills, 1985). For instance, Constantinople, Cornelius, and Gray (1988) assessed teacher and student gender, type of curriculum, and time of semester in first and second year courses at Vassar College, and found that male students spoke up more frequently in female-led classes, initiated more classroom discussion, and participated more in class overall. Studies using students from significantly sized universities or college students in large samples show similar sex differences (Brady & Eisler, 1995; Crawford & MacLeod, 1990).

Researchers now admit that classroom climate both affects and is affected by more than student gender. In fact, the reported student gender differences may be artifactually produced by variables that have largely gone unstudied, such as teacher gender and the student-teacher interaction. Brady and Eisler (1995, p. 14) write that "the mere presence of a male instructor, not his behavior, may have contributed to an intimidating environment for females." Given that female instructors remain under-represented in college faculty, and given that the ratio of male to female instructors is not consistent across academic discipline, students' classroom experience may be dictated by the instructor's gender, the course material instructed, and the students' gender-related expectations (Deaux & Major, 1987). "Because students expect female professors to be more personal, supportive and motherly than male professors, it is possible that the students' gender-role expectations influenced their relative amount of participation" (Brady & Eisler, 1995, p. 14).

In short, if the presence of female instructors activate more gender stereotypes (and unique evaluations) among students in disciplines occupied less by female faculty (e.g., sciences) compared to those occupied more (e.g., social sciences), then evidence of a chilly college climate would be revealed through differential instructor evaluations, a critical tool utilized by institutions to help in decisions concerning personnel retention, promotion, tenure, and salary increases (Cashin & Downey, 1992; Divoky & Rothermel, 1988). As a result, the favorability of professors could interact among many variables, such as student sex, professor sex, and discipline. For example, Divoky and Rothermel (1988) assessed students' instructor ratings based on the importance students held for the dimensions of teaching effectiveness (e.g., delivery, depth of knowledge, interpersonal skills). Results showed that: (a) instructor delivery was important to students in non-major courses compared to major courses; (b) instructor depth of knowledge was important for students in major-elective courses than either non-major elective or required courses; and (c) instructor interpersonal skills were important for students in major required courses than students in major elective courses. The authors concluded (p. 45) that "instructors should not be compared by their mean ratings on each item on an evaluation form but rather a weighted average rating could be calculated for an instructor in each type of course he or she teaches."

Overall, one may ask whether student bias has created unequal evaluation criteria for male and female faculty. Whereas most research indicates that female professors receive higher evaluations than male professors (Feldman, 1993; Freeman, 1994; Hom, DeNisi, Kinicki, & Bannister, 1982; Marsh, 1980), these studies often fail to account for many important intervening variables such as student sex, professor rank and discipline, and student personality traits. When researchers account for the influence of these variables, results show female faculty receiving lower ratings than male faculty, especially from males students (Basow & Silberg, 1987; Murray, Rushton, & Paunonen, 1990).

Recently, Basow (1995) examined gender bias in faculty evaluations based on students' assessments in three disciplines (humanities, natural, and social sciences) for each of four years at Lafayette College. In addition

to professor sex, student sex, and discipline, Basow assessed the hour the class met, professor rank and teaching experience, and student year, grade point average, and expected final grade. Across the four years, results consistently showed that male faculty were rated similarly by their female and male students, regardless of discipline affiliation, whereas female faculty were rated highest by female humanities students but lowest by male social science students.

Whereas Basow delineated many of the influences on instructor evaluation, there are several concerns with the research that merit discussion. First, by using over 2000 subjects in a multivariate analysis, Basow achieved an astoundingly high level of statistical power, which yielded many significant but largely trivial effects. Second, students were sampled from a small liberal arts college in the Northeastern United States, which only a few years prior began to admit female students. This bias does not invalidate the findings, but does invite a similar investigation at a larger, more typical North American institution. As Brady and Eisler (1995) recommend, a large sample of students and teachers taken from multiple departments would help to increase the representativeness and generalizability of the sample drawn. Finally, because Basow conducted the research in the United States, it warrants a similar (although modified) investigation in Canada for the purposes of comparison, standardization, and validation. Due to the typical multicultural student profile at Canadian universities, faculty evaluations may also be influenced by social characteristics such as ethnicity, disability, Aboriginal status (see Chambers, Lewis, & Kerezsi, 1995; D'Augelli & Hershberger, 1993) that have remained largely unstudied.

## **Present Study**

To address the issues outlined above, the present study examines several basic questions concerning students' differential evaluation of faculty instruction as affected by student gender, instructor gender, discipline, and a variety of demographic and social variables. Is faculty evaluation significantly and noticeably influenced by variables such as faculty gender, student gender, and/or discipline? Is such an influence affected by student and faculty social variables such as minority status?

As a modified replication, the present study evaluated Canadian university professors from the college of arts and sciences using the same measurement instrument as Basow. Additionally, other sources of bias such as Aboriginal, disability, and minority status, were assessed to draw out potential covariates. Are the interaction effects reported by Basow (1995) replicable in a Canadian sample, and if so, do the effects remain small enough to be practically unimportant?

Past research has been mixed in outlining the relationship between faculty and student variables in the evaluation of course instruction (Basow, 1995; Feldman, 1993; Murray et al., 1990). Whereas Basow (1995) found significant effects of faculty, student, and discipline variables on evaluations, the findings are qualified by significant but practically unimportant results and investigation at an atypical institution. By modifying Basow's original work at a Canadian institution, results from the present study should: (1) help researchers understand the extent of student bias in faculty evaluations; (2) help administrators better evaluate their teaching faculty; (3) better guide teaching faculty to improve their quality of instruction to students; and (4) evaluate the degree of student and faculty sensitivity to critical social issues.

## METHOD

### Participants

**Professor respondents.** Overall, the profile of the 293 tenured and tenure-track professors in the College of Arts and Sciences at the University of Saskatchewan appears as follows: 244 males (83%) and 49 females (17%); 164 full professors (56%), 95 associate professors (32%), and 34 assistant professors (12%). By discipline, there are 78 male and 28 female professors in Fine Arts/Humanities, 64 male and 13 female professors in Social Sciences, and 102 male and 8 female professors in Sciences. Using these characteristics, classes were stratified according to discipline and year to maximize both sample-to-population similarity and generalization of results across the College. Furthermore, in order to sample across different class sizes and still recruit between 800 and 1200 student respondents, 32 professors (approximately 11%)

were randomly selected and both they and their students were invited to participate in the investigation. Of those 32 professors approached, 28 (8 females, 20 males) agreed to participate (1 female and 3 males refused), resulting in 910 student respondents. Overall, the characteristics of this sample demonstrated adequate similarity to the overall profile in the College of Arts and Sciences.

**Student respondents.** There were 910 undergraduate students (399 males, 509 females, and 2 unspecified) enrolled in the College of Arts and Science at the University of Saskatchewan who participated in this study. Half of the sample ( $n = 455$ ) was 19 to 21 years old; 151 students (17%) were younger than 19 years, 177 students (20%) were 22 to 24 years old, and the remainder ( $n = 125$ , 14%) was older than 24 years. Ninety-six students (11%) were members of a visible minority, 25 students (3%) were Aboriginal, and 43 students (5%) had a disability.

There were 302 (33%) first-year students, 223 (25%) second-year students, 166 (18%) third-year students, and 135 (15%) fourth-year students. The remaining 84 students (9%) either did not indicate their year or were beyond their fourth year in university. Students were enrolled in Science ( $n = 335$ , 37%), Fine Arts and Humanities ( $n = 205$ , 23%), or Social Science ( $n = 192$ , 21%) programs; 178 students (20%) were in enrolled in "Other" programs (e.g., combined majors).

Students reported their current and expected final marks in the course they were evaluating for this study. For their current marks, 240 students (26%) reported an average between 80% and 89%, 315 students (35%) reported an average between 70% and 79%, and 205 students (23%) reported an average between 60% and 69%. The remaining 141 students (16%) reported averages that were either above 89% or below 60%. Nine students did not report their current average mark.

Students' expected final course marks were distributed similarly: 303 students (33%) expected a final mark between 80% and 89%, 354 students (39%) expected a final mark between 70% and 79%, and 143 students (16%) expected a final mark between 60% and 69%. Seventy-five students (8%) expected a final mark that was either above 89% or below 60%, and 35 students (4%) did not answer this question.

## Measures

**Students' Teaching Needs Questionnaire.** This three-part scale, based on 19 items developed by Basow (1995), asks students to evaluate their professor's teaching style in a given course. Specifically, students indicate the extent to which they believe their professor is enthusiastic, fair, helpful, knowledgeable, organized, and sensitive to their needs. Furthermore, students rate to what extent the professor treats them with respect, stimulates thinking, expresses ideas well, avoids unnecessary repetition, speaks in an appropriate manner, provides clear explanations and objectives, good feedback and fair grading, and assigns appropriate readings and tests/papers. Finally, students provide an overall evaluation of both course and professor. The present scale includes three additional items on professors' sensitivity to targeted student groups such as visible minorities or disabled students (e.g., "Your instructor is sensitive to the needs of students with disabilities").

In the first section (16 items), students rate their professor on helpfulness, sensitivity to students' learning needs, teaching style, and teaching effectiveness using a 5-point Likert scale format (1 = Strongly Disagree, 5 = Strongly Agree). In the second section (6 items), students rate the overall course structure, requirements, and objectives using a 5-point Likert scale format (1 = Poor, 5 = Excellent). In the third and final section (9 items), students indicate their gender, age, designated group status (e.g., visible minority, Aboriginal, disabled), academic major, years in university, and current and expected final marks in the course.

**Professor's Questionnaire.** Professors in each of these courses provided information about their course and about their teaching experience on a six-item questionnaire. The questions asked for professor's gender and years of teaching, as well as the number of hours the professor met with students outside of class, number of students in the class, when the class was held (Morning, Afternoon, or Evening), and how often the class met per week.

For the 28 courses used in this study, the 8 female and 20 male professors had an average of 15.3 years ( $SD = 11.5$ ; range from 3 to 40 years) teaching experience. The average class size was 59 students



( $SD = 76$ ), ranging from 8 to 330 students. Twenty courses were held in the morning, seven in the afternoon, and one in the evening. About half of the courses ( $n = 16$ , 57.1%) met three times a week, four courses met twice a week, and eight courses met once a week. Professors reportedly met with students an average of 4.2 hours per week ( $SD = 4.0$ ) outside of class time.

## **Procedure**

Courses were randomly selected from three discipline areas (Fine Arts and Humanities, Social Sciences, and Sciences) in the College of Arts and Science. The researchers contacted professors of the selected courses to ask for their participation. For those professors who agreed to participate in the study, a research assistant administered a consent form and the *Students' Teaching Needs Questionnaire* to each student during class time. While students were completing their questionnaire package, their professor completed a consent form and filled out the *Professor's Questionnaire*. Both the student and professor consent form emphasized the confidentiality of responses and the voluntary nature of the participation. Professors and students were debriefed and thanked for their participation.

## **RESULTS**

### **Evaluation Scale Structure**

Using principal components analysis with oblique rotation, the 910 student responses on the 22-item questionnaire were reduced to four factors whose eigenvalues before rotation exceeded unity. The solution accounted for approximately 64% of the available factor space. The first factor, labelled *Course Structure* (accounting for 42% of the factor space; eigenvalue = 9.22), consisted of 6 items of students' impressions of the course tests, readings, objectives, etc. The second factor, labelled *Student Needs* (10%, 2.18), consisted of 5 items of students' impressions of how sensitive the professor was to their own academic experience (e.g., respect, sensitivity, helpfulness). The third factor, labelled *Teaching Quality* (7%, 1.56), consisted of 6 items of students' impressions of the

professor's ability to stimulate thinking, present clear and organized ideas, etc. The fourth factor, labelled *Treatment of Target Groups* (5%, 1.19), consisted of 3 items tapping students' impressions of their professors' sensitivity to the issues of minority groups, women's issues, and the disabled.

Based on these results, four subscales were constructed as the sum of their component items (e.g., Target Group Treatment Subscale equalled the sum of the three component items). Two items were excluded because they did not load significantly on any of the four factors. The four subscales correlated significantly with one another, ranging from .32 for Course Structure and Target Group Treatment, to .67 for Student Needs and Teaching Quality. In general, the overall evaluation scale and component subscales indicated that students gave moderately positive ratings to their courses and professors (see Table 1 for means, standard deviations, and internal consistency reliability estimates).

Because of the sizeable number of participants ( $N = 910$ ) used in this correlational analysis, even modest correlations were statistically significant. Therefore, only correlations that accounted for greater than 10% of the variance ( $r > .32$ ) were considered important and interpretable. None of the correlations met this criterion. In fact, the relation between the subscales and each of Aboriginal, minority, and disabled status was not significant.

### **Tests for Gender and Discipline Differences**

The evaluation scale and subscales (Overall Evaluation, Course Structure, Student Needs, Teaching Quality, Treatment of Target Groups) served as dependent variables in a series of  $2 \times 2 \times 3$  analyses of variance (ANOVAs), with Student Gender (male vs. female), Professor Gender (male vs. female), and Discipline (Science, Social Science, Fine Arts/Humanities) as the independent variables (see Table 2 for means, standard deviations, and group sizes; see Table 3 for an ANOVA summary). Because group sizes were not equal (i.e., a nonorthogonal design), Type III unique sums of squares were utilized to assess all main effects and interactions independent of each another. In addition, given the large sample size and comparably high statistical power, the  $\omega^2$

Table 1  
**Means, Standard Deviations, Reliability, and Correlations of Evaluation Subscales with Student Variables**

Evaluation Scale (items)	Mean	SD	$\alpha$	Range	Student Correlates		
					Age	University Year	Current Expected Grade
Student needs (6)	19.67	4.25	.87	6-30	.09	.04	.15
Teaching Quality (5)	24.12	4.63	.81	5-25	.10	.03	.11
Target Group Treatment (6)	10.73	2.34	.87	6-30	.05	-.02	.07
Course Structure (3)	21.86	5.16	.88	3-15	.06	.01	.29
Total Evaluation (20)	76.38	13.26	.93	20-100	.09	.02	.21

Note:  $N = 910$

**Table 2**  
**Evaluations Subscale Means (Standard Deviation) by Student Gender, Professor Gender, and Discipline**

	Student Gender		Discipline			Professor Gender	
	Male ( <i>n</i> = 399)	Female ( <i>n</i> = 507)	Social Sciences ( <i>n</i> = 335)	Sciences ( <i>n</i> = 205)	Fine Arts/ Humanities ( <i>n</i> = 192)	Male ( <i>n</i> = 20)	Female ( <i>n</i> = 8)
Evaluation Scale							
Student needs	19.20 (4.50)	20.03 (4.02)	19.26 (4.49)	20.17 (3.72)	20.31 (3.86)	19.93 (2.41)	21.08 (1.44)
Teaching Quality	23.63 (4.92)	24.49 (4.35)	23.37 (4.79)	24.96 (4.43)	24.98 (4.18)	24.40 (2.49)	25.68 (1.70)
Target Group Treatment	10.41 (2.28)	10.96 (2.36)	10.33 (2.21)	10.97 (2.33)	11.19 (2.48)	10.67 (1.03)	11.86 (1.30)
Course Structure	21.67 (5.40)	21.98 (4.96)	21.33 (5.60)	22.56 (4.45)	22.50 (4.80)	21.94 (2.67)	22.72 (1.29)
Overall Evaluation	74.85 (13.97)	77.36 (12.62)	74.17 (13.84)	78.49 (11.77)	78.98 (12.16)	76.94 (7.53)	81.35 (4.97)

**Table 3**  
**Analysis of Variance Summary Table by Subscale and Effect**

<b>Subscale</b>						
Main Effect/Interaction						
Simple Effects	<i>F</i>	dfs	<i>p</i>	MSE	$\omega^2$	
<b>Overall</b>						
Discipline	2.82	2,700	.060	164.80	.005	
Discipline x Student Gender	3.26	2,700	.039	164.80	.006	
Student Gender (Science only)	3.98	1,324	.047	190.79	.013	
<b>Student Needs</b>						
Student Gender	4.53	1,715	.034	16.90	.005	
<b>Teaching Quality</b>						
Discipline	4.09	2,717	.017	20.42	.002	
Student Gender	3.57	1,717	.059	20.42	.003	
Discipline x Student Gender	4.13	2,717	.017	20.42	.008	
Discipline (Males only)	8.31	2,328	.001	23.11	.042	
Discipline (Females only)	3.69	2,325	.026	18.13	.013	
<b>Targets</b>						
Discipline	5.62	2,705	.004	27.89	.013	
Student Gender	5.25	1,705	.022	27.89	.006	
Professor Gender	4.19	1,705	.041	27.89	.013	

Note: Nonsignificant main effects and interactions are not listed.

statistic estimates the proportion of variance explained by the factors (Howell, 1996). This statistic is useful in identifying significant but trivial effects, where values below .03 are deemed unimportant.

The ANOVA of students' overall professor evaluations showed a marginal main effect for Discipline; but a significant Student Gender x Discipline interaction. Mean inspections showed that Science students gave a slightly lower overall evaluation compared to Social Science and Fine Arts/Humanities students. Simple effects tests of the interaction showed that for Science students, males gave significantly lower overall evaluations than female students ( $M_s = 72.98$  and  $76.02$ , respectively). There were no significant differences by Student Gender for Fine Arts/Humanities or Social Sciences.

Although there were no significant differences in Course Structure, there were significant differences in Student Needs by Student Gender, whereby female students believed their professors were more sensitive to their needs than male students. For Teaching Quality, there was a marginal main effect by Student Gender, a significant main effect by Discipline; and a significant Student Gender x Discipline interaction. Mean inspections revealed that female students gave slightly higher ratings than male students. As a test of the pairwise differences among the three levels of Discipline, the REGWQ multiple comparison procedure was selected because it offered the most powerful test of mean differences when the number of groups is few, while still controlling the familywise error rate (Howell, 1996, p. 381). In addition, the Games-Howell correction method (Toothaker & Miller, 1996) was used to account for unequal group sizes. Results showed that Science students gave significantly lower ratings on teaching style than both Social Science and Fine Arts/Humanities students, whose ratings did not differ. To assess the nature of the interaction, a simple effects test of Discipline (controlling for Student Gender) showed a significant effect for male students, whereby male Science students were more critical of their professor's teaching style than male Fine Arts/Humanities students ( $M_s = 22.97$  and  $25.69$ ). The test was also significant for female students, whereby female Science students were more critical than female Social Science students ( $M_s = 23.96$  and  $25.39$ ).

An ANOVA for Treatment of Target Groups showed significant main effects for Student Gender, Professor Gender, and Discipline. Inspection of the means revealed that female students believed their professors were more sensitive to target group issues than male students. In addition, female professors were rated as more sensitive to target group issues than male professors. REGWQ tests showed that professors were judged to be less sensitive to target group needs by students in the Sciences than by students in Social Sciences or Fine Arts/Humanities.

### **Class Analyses**

Since there was a considerable range in class size (8 to 330 students), the responses from students in the larger classes would have biased the findings concerning individual courses or professors; therefore, *mean evaluation scores* for each class were calculated and used in analyses with course and professor variables to give equal weighting to all courses. Table 4 shows the mean class evaluation scores and the correlation matrix for the evaluation scales with course variables (class size, number of times the class meets per week, time of day the class meets) and with professor variables (years of teaching, the number of hours the professor meets with students outside of class, professor gender).

Results showed that professor sensitivity to Student Needs was correlated positively with time of day in which the class was held (morning versus afternoon/evening). This finding may be an indicator of students' reactions to class size since morning classes tended to be larger classes, and Student Needs were somewhat negatively correlated with class size. Course Structure correlated positively with professor's years of teaching, and Teaching Quality correlated positively with the number of hours a professor met with students outside of class. In addition, professor gender was correlated significantly with Target Group Treatment, and correlated somewhat with Overall Evaluation (i.e., students rated female professors more positively than male professors). Lastly, Overall Evaluation was somewhat negatively correlated with class size, but somewhat positively correlated with the number of hours a professor met with students outside of class.

Table 4  
**Correlations of Evaluation Subscales with Professor and Course Variables**

Evaluation Scale	Mean	SD	Course			Professor		
			Class Size	Meetings per week	Time of day	Gender	Years of Teaching	Outside Class Hours
Student needs	20.26	2.21	-.25	-.23	.38*	.24	.44*	.29
Teaching Quality	24.77	2.34	-.22	-.11	.03	.25	.26	.49*
Target Group Treatment	11.01	1.22	-.22	-.38*	.06	.45*	.28	.30
Course Structure	22.17	2.36	-.09	-.03	-.05	.15	.38*	.25
Total	78.20	7.10	-.22	-.19	.02	.29	.40*	.38*

Note:  $N = 28$ ; \*  $p < .05$

Correlations by time of day (morning vs. afternoon/evening) and gender (male, female) are point-biserial; the remainder are Pearson.



## DISCUSSION

Both the strengths and weaknesses of the present study should be considered when interpreting the results. The study expanded upon Basow's (1995) work by using a more generalizable sample of students, and by including additional variables that may be relevant to students' evaluations of their professors (e.g., student's designated group status, the number of hours a professor met with students outside class). However, only cautious interpretations can be drawn from the results because (as in Basow's study) the student and professor variables still only explained a small amount of the variance in evaluations scores, and there were too few female professors to adequately test differences in evaluation scores based on professor gender.

Results showed that students in this sample gave moderately positive ratings to many qualities of their professors' teaching, including style, sensitivity to student needs, and aspects of the course itself. Although one might anticipate variation in student ratings based on student age, year in university, or target group designation (e.g., female, Aboriginal, disability), the only notable effect occurred with respect to course structure, which was moderately related to students' current and expected final grades (i.e., students performing better academically more likely believed their course had clear objectives, relevant readings, and fair examinations).

However, students' gender and field of study were closely related to how students evaluated their professor and course. Female students perceived professors to be more sensitive to their academic needs and to the needs of target groups than male students. In addition, female students felt that their professors were better teachers than did male students. Specifically, female Social Science students and male Fine Arts/Humanities students were more positive than either female or male Science students about their professor's teaching style.

In general, Science students were more critical of their professor's teaching style and sensitivity to target group needs than were either Social Science or Fine Arts/Humanities students. These findings are consistent with Holdaway and Kelloway (1987), who found that first-year

Science students were significantly less satisfied with the university experience than first-year Arts students. Similarly, Worth, Crombie, and Rinholm (1991) found that Arts students reported their professors engaged in more personalized classroom behaviour (e.g., called on students by name) and had more interactive exchanges with students than Science students reported experiencing.

It may be that there is little opportunity for discussion of target group issues in Science classes when compared to Social Sciences or Fine Arts/Humanities classes. In fact, courses in the latter two disciplines often include presentations of women's issues, Native issues, and topics of racism or sexism as part of their curriculum. Thus, students may perceive that their Social Science and Humanities professors are more sensitive to such issues than their Science professors. Furthermore, the Sciences are viewed as traditionally male-dominated fields and this may further contribute to students' less positive reactions to their Science professors' sensitivity and teaching.

Characteristics of professors and their courses were also associated with students' evaluations of their professors. For instance, professors of larger classes were more likely to be rated less positively than professors of smaller classes, especially in terms of meeting the academic needs of students. Crawford and MacLeod (1990) also found that class size affected many aspects of classroom climate such as participation rates and perceptions of professors' behaviour. In addition, professors with more teaching experience, and who scheduled more office hours with students, received higher evaluations of their teaching ability and course structure. Lastly, female professors were rated as more sensitive to target group issues than their male counterparts.

### **Implications and Directions for Future Research**

As in Basow (1995), the variables included in the present study explained only a minimal amount of the variance in students' evaluation scores; however, the findings of both this and Basow's (1995) studies indicate there are multiple influences on students' ratings of their professors, operating either as main effects or as interactions at any one time. That is, the interaction of students' gender, expected and current and

expected final grades, area of study, and additional (as yet undetermined) factors may play a significant role in students' evaluation of their professors and courses. It should be noted that although these variables have little to do with a professor's actual teaching ability, they nonetheless may be affecting students' reactions to their professors.

Previous research (e.g., Alexitch, 1994; Divoky & Rothermel, 1988; Morstain, 1977; Remigio & Page, 1991) has indicated that students' values, motivational orientation, and their reasons for choosing courses or for pursuing a post-secondary education may affect students' expectations and satisfaction with their courses, programs, and professors. For instance, some researchers (Morstain, 1977; Remigio & Page, 1991) have noted that the congruence between students' values and views about the purpose of education with those of faculty may affect students' level of satisfaction with their university education. In addition, there is evidence indicating that students' academic aspirations and satisfaction may be significantly influenced by informal (outside class) rather than formal (in-class) faculty-student contact (Lampert, 1993; Pascarella, 1984; Pascarella, Terenzini, & Hibel, 1978; Theophilides, Terenzini, & Lorang, 1984).

The campus climate for students is largely comprised of many of these aspects: students' perceptions of higher education, academic expectations, programmatic demands and restrictions, and their degree of satisfaction with the people who make up the campus community (Baird, 1990; Pascarella, 1984). Since some or all of these variables may affect students' ratings of professors, future research examining students' evaluations of their professors should endeavour to incorporate additional student, programmatic, and campus-related variables. D'Apollonia and Abrami (1997) advised that while students' ratings of professors were moderately valid, they alone were not sufficient to gauge teaching ability in professors. Therefore, administrators and faculty should consider administrative and programmatic variables (e.g., class size, program requirements) and both professors' and students' gender when using teaching evaluations for tenure and promotion decisions. This focus would effectively increase the validity of students' ratings of professors, and thereby provide a more

accurate reflection of a professor's teaching. More importantly, students should benefit from higher quality teaching offered by professors who receive more constructive teaching evaluations. ✨

## References

- Alexitch, L.R. (1994). *Undergraduate student expectations and perceptions of a university education in the 1990s*. Unpublished doctoral dissertation, University of Windsor, ON.
- Baird, L.L. (1990). Campus climate: Using surveys for policy-making and understanding. *New Directions for Institutional Research*, 68, 35–45.
- Basow, S.A. (1995). Student evaluations of college professors: When gender matters. *Journal of Educational Psychology*, 87, 656–665.
- Basow, S.A., & Silberg, N.T. (1987). Student evaluation of college professors: Are female and male professors rated differently? *Journal of Educational Psychology*, 74, 170–179.
- Brady, K.L., & Eisler, R.M. (1995). Gender bias in the college classroom: A critical review of the literature and implications for future research. *Journal of Research and Development in Education*, 29, 9–19.
- Carelli, A.O. (1988). *Sex equity in education: Readings and strategies*. Springfield, IL: Charles C. Thomas.
- Cashin, W.E., & Downey, R.G. (1992). Using global student rating items for summative evaluation. *Journal of Educational Psychology*, 84, 563–572.
- Chambers, T., Lewis, J., & Kerezsi, P. (1995). African American faculty and White American students: Cross-cultural pedagogy in counselor preparation programs. *Counseling Psychologist*, 23, 43–62.
- Constantinople, A., Cornelius, R., & Gray, J. (1988). The chilly climate: Fact or artifact? *Journal of Higher Education*, 59, 527–550.
- Crawford, M., & Macleod, M. (1990). Gender in the college classroom: An assessment of the 'chilly climate' for women. *Sex Roles*, 23, 101–122.
- D'Apollonia, S., & Abrami, P.C. (1997). Navigating student ratings of instruction. *American Psychologist*, 52, 1198–1208.
- D'Augelli, A.R., & Hershberger, S.L. (1993). African American undergraduates on a predominantly white campus: Academic factors, social networks, and campus climate. *Journal of Negro Education*, 62, 67–81.
- Deaux, K., & Major, B. (1987). Putting gender into context: An interactive model of gender-related behavior. *Psychological Review*, 94, 369–389.

Divoky, J.J., & Rothermel, M.A. (1988). Student perceptions of the relative importance of dimensions of teaching performance across type of class. *Educational Research Quarterly*, 12, 40–45.

Feldman, K. (1993). College students' views of male and female college teachers: Part II — Evidence from students' evaluations of their classroom teachers. *Research in Higher Education*, 34, 151–211.

Freeman, H.R. (1994). Student evaluations of college instructors: Effects of type of course taught, instructor gender and gender role, and student gender. *Journal of Educational Psychology*, 86, 627–630.

Frale, D.E.S. (1989). Sex typing and gender ideology: Two facets of the individual's gender psychology that go together. *Journal of Personality and Social Psychology*, 56, 95–108.

Harvey, G., & Hergert, L.F. (1986). Strategies for achieving sex equity in education. *Theory Into Practice*, 25, 290–299.

Heller, J.F., Puff, C.R., & Mills, C.J. (1985). Assessment of the chilly college climate for women. *Journal of Higher Education*, 56, 446–461.

Holdaway, E.A., & Kelloway, K.R. (1987). First year at university: perceptions and experiences of students. *The Canadian Journal of Higher Education*, 17, 47–63.

Hom, P.W., DeNisi, A.S., Kinicki, A.J., & Bannister, B. (1982). Effectiveness of performance feedback from behaviourally anchored ratings scales. *Journal of Applied Psychology*, 67, 568–576.

Howell, D.C. (1996). *Statistical methods for psychology* (4th ed.). Toronto, ON: ITP Nelson.

Lampert, M.A. (1993). Student-faculty informal interaction and the effect on college student outcomes: A review of the literature. *Adolescence*, 28, 971–990.

Marsh, H.W. (1980). The influence of student, course, and instructor characteristics in evaluations of university teaching. *American Educational Research Journal*, 14, 441–447.

Marsh, H.W., & Roche, L.A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52, 1187–1197.

McKeachie, W.J. (1997). Student ratings: The validity of use. *American Psychologist*, 52, 1218–1225.

Morstain, B.R. (1977). An analysis of students' satisfaction with their academic program. *Journal of Higher Education*, 48, 1–16.

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.

Pascarella, E. (1984). College environmental influences on students' educational aspirations. *Journal of Higher Education, 55*, 751–771.

Pascarella, E., Terenzini, P., & Hibel, J. (1978). Student-faculty interactional settings and their relationship to academic performance. *Journal of Higher Education, 49*, 450–463.

Remigio, J., & Page, S. (1991). Value orientations in Canadian university undergraduates. *Journal of Psychology and Behavioural Sciences, 6*, 160–166.

Theophilides, C., Terenzini, P.T., & Lorang, W. (1984). Relation between freshman year experience and perceived importance of four major educational goals. *Research in Higher Education, 20*, 235–253.

Toothaker, L.E., & Miller, L. (1996). *Introductory statistics for the behavioral sciences* (2nd ed.). New York, NY: Brooks/Cole.

Williams, D. (1990). Is the post-secondary classroom a chilly one for women? A review of the literature. *The Canadian Journal of Higher Education, 23*, 29–42.

Worth, D.A., Crombie, G., & Rinholm, J. (1991, June). *Students' perceptions: Gender differences in the university classroom experience*. Paper presented at the annual meeting of the Canadian Psychological Association, Calgary, AB.