

DISSERTATION

FACULTY POWER AND STUDENT INTEGRITY:
STUDENT RATINGS OF TEACHING IN A COMMUNITY COLLEGE

Submitted by

Craig Hadden

School of Education

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2004

UMI Number: 3143829

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform 3143829

Copyright 2004 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

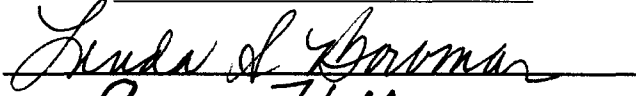
ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

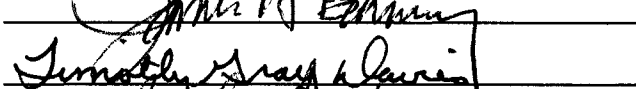
COLORADO STATE UNIVERSITY

June 3, 2004

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED
UNDER OUR SUPERVISION BY CRAIG HADDEN ENTITLED "FACULTY
POWER AND STUDENT INTEGRITY: STUDENT RATINGS OF TEACHING IN
A COMMUNITY COLLEGE" BE ACCEPTED AS FULFILLING IN PART
REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work





Adviser

Director

ABSTRACT OF DISSERTATION

FACULTY POWER AND STUDENT INTEGRITY:
STUDENT RATINGS OF TEACHING IN A COMMUNITY COLLEGE

Despite numerous studies aimed at confirming the reliability, validity, and generalizability of student ratings of teaching, most of them concern the instrument itself. These studies are quantitative or meta-analyses rather than qualitative. The present grounded theory study concerns student candor and sincerity when completing these ratings. From the interviews with 12 students, the core category of power emerged. Through remarks instructors make about ratings as well as through their communication with students, their general attitude and personality traits, their treatment of students, and their pedagogical skills and practices, instructors have the power to create climates of fear or trust. Student candor and sincerity depend to a great extent on the level of fear or trust they feel, thus impacting the trustworthiness of instruments that may otherwise be reliable and valid.

Charles Craig Hadden
School of Education
Colorado State University
Fort Collins, CO 80523
Summer 2004

Acknowledgements

Thanks to my mother for reading to me every night as a child, thus instilling in me a love of language. Thanks to my father for his constant encouragement to pursue my education to the limit. Thanks to my committee members for their patience in reading such a long dissertation. Thanks to Dr. Linda Bowman, whose wisdom kept me focused on my goal and who saw potential in me that I did not necessarily see in myself.

Special thanks to Dr. Timothy G. Davies, my advisor, my mentor, and my friend. His guidance and support have been indispensable in helping me through five years of study.

TABLE OF CONTENTS

CHAPTER ONE – INTRODUCTION.....	1
Background.....	1
Statement of the Research Problem.....	7
Research Questions.....	8
Definition of Terms.....	9
Significance of the Study	10
Researcher’s Perspective	11
CHAPTER TWO – REVIEW OF THE LITERATURE.....	14
Studies Concerning Consequential Validity.....	19
Studies Concerning Conceptual Structure.....	22
Studies Concerning Convergent Validity.....	29
Studies Concerning Discriminant Validity.....	35
Discriminant Validity (Bias) Studies Related to Course Variables.....	40
Non-Bias Factors.....	40
Course Level.....	40
Class Size.....	41
Workload/Difficulty.....	44
Administration and Stated Purpose of the Ratings.....	45
Academic Discipline.....	47
Discriminant Validity (Bias) Studies Related to Student Variables.....	47
Non-Bias Factors.....	48
Prior Subject Interest.....	48
Reason for Taking the Course	49
Motivation.....	50
Student Level.....	51
Instructor Gender.....	52
Consumerism.....	57
Feelings of Indebtedness.....	59
Retribution.....	59
Fear of Retaliation.....	60
Apathy.....	61
Instructor Reputation.....	62
Halo Effect.....	63
Discriminant Validity (Bias) Studies Related to Instructor Variables.....	71
Non-Bias Factors.....	72
Rank.....	72
Years of Experience.....	74
Type of Instruction.....	75
Grading Leniency/Expected Grade.....	76
Instructor Personality.....	84
Enthusiasm (the “Dr. Fox Effect”).....	86
Instructor Likeability (Rapport).....	91
Summary of the Research.....	94
CHAPTER THREE – METHOD	98

Research Design.....	98
Participants and Site.....	100
Data Collection, Instruments, and Procedure.....	103
Data Analysis.....	105
CHAPTER FOUR – RESULTS.....	107
The Paradox of Candor and Sincerity.....	110
The Phenomenon of Untrustworthy Ratings.....	111
Fear – The Key Factor Underlying Lack of Candor and Sincerity on SRTs.....	119
Fear of the Unknown.....	121
Fear Related to the Instructor.....	126
Fear Related to the Administration.....	134
Fear Related to Being Perceived as Biased.....	140
Fear Related to the Students.....	141
Instructor Behavior and a Climate of Fear.....	143
Actions in Relation to the Ratings.....	144
Communication Skills.....	146
Attitude and Personality Traits.....	148
Treatment of Students.....	151
Pedagogical Skills and Practices.....	152
Administration, the System, and a Culture of Fear.....	154
The Phenomenon of Trustworthy Ratings.....	155
Trust – The Key Factor Underlying Candor and Sincerity on SRTs.....	161
Preconceptions about Responsibilities and Factors That May Impact Ratings.....	161
Knowledge of, Beliefs, and Attitudes toward the Rating Process.....	164
Student Perceptions of Instructors.....	166
Positive Instructor Behavior and the Culture of Trust.....	168
Actions in Relation to the Ratings.....	168
Communication Skills.....	174
Attitude and Personality Traits.....	176
Treatment of Students.....	179
Pedagogical Skills and Practices.....	181
The Core Category: Abuse and Use of Power.....	187
CHAPTER FIVE – DISCUSSION.....	209
Summary of the Findings That Negate or Partially Negate the Literature.....	213
Summary of the Findings That Support the Literature.....	227
Findings Related to Consequential Validity.....	227
Findings Related to Students and Discriminant Validity.....	230
Course-Related Biases.....	230
Purpose of the Ratings.....	231
Class Size in Relation to Student Contact and Power.....	233
Student-Related Biases.....	235
Prior Subject Interest and Reason for Taking the Course.....	235
Consumerism.....	236
Feelings of Indebtedness.....	237
Apathy.....	237

Reputation.....	238
Halo Effect.....	239
Instructor-Related Biases.....	243
Personality/Likeability.....	243
Enthusiasm.....	245
Findings Related to Characteristics Associated with Superior College	
Instructors.....	248
Findings Related to Empowerment and Active Learning.....	250
Discussion of the Findings That Extend Beyond the Literature.....	251
Future Research.....	278
Recommendations Based on the Findings.....	281
Final Thoughts.....	285
Epilogue.....	286
REFERENCES.....	292
APPENDIX A.....	302

CHAPTER ONE

INTRODUCTION

Background

Perhaps as a result of reports such as *A Nation at Risk* in the 1980s, the public in general as well as educators and accrediting agencies in particular began to attend to the issue of the assessment of teaching and learning. Although some may have assumed this issue of assessment was just another educational fad, now over 20 years later state and national legislatures, accrediting agencies, and colleges and universities concern themselves with setting performance standards, creating and implementing assessment plans, and generally demonstrating to a variety of stakeholders that students graduating from higher education institutions are leaving with greater knowledge and skills than when they entered. As evidence of this fact, one can examine Colorado, in which the Colorado Commission on Higher Education (CCHE) and the state legislature have established a Quality Indicator System (QIS), setting a variety of performance indicators for all institutions of higher education. Furthermore, the North Central Association (NCA), which accredits most of the institutions of higher learning in Colorado (and all of the state-funded ones), mandates assessment plans along with evidence of implementation and results as part of the reaccreditation process, and this association is now allowing institutions to use the Academic Quality Improvement Program (AQUIP) model of assessment as well. The Western Association of Schools and Colleges, under whose auspices California institutions fall, has

now become the last of the six regional accrediting associations in the United States to require outcomes assessment as part of reaccreditation. Within most institutions, one of the tools common to all and often used as part of the general assessment process for faculty is the faculty/course rating instrument with which students evaluate or rate instructors and courses.

Using student ratings of professors with the goal of improving teaching is nothing new. According to Cruse (1987), this practice has been recurring since the early 1900s. Besides providing diagnostic information to faculty members about their teaching, some institutions also use these ratings as a measure of teaching effectiveness in tenure/promotion decisions, as a means of providing information for student use in selecting courses and instructors (when the ratings are published), and as an outcome on a process description for research or teaching (Cohen, 1980; El-Hassan, 1995; Freedman & Stumpf, 1978; Jeffreys, M.R., Massoni, M., O'Donnell, M., & Smolaka, I., 1997; Marsh 1987; Miller, 1988; Nassar & Fresko, 2002; Ory, 1990; Pike, 1998). Marsh (1987) includes a fifth purpose, that of providing a measure of the quality of the course for course improvement and curriculum development, but concludes student ratings of teaching (SRTs) are probably not very useful for these aims, and Miller (1988) includes being responsible to society. Although student ratings are generally not included as a factor that accrediting agencies and other external stakeholders consider, it is possible that student ratings along with other evaluations of instructors may one day become a routine part of the overall assessment process of institutions. Even if evaluations are not included as part of an overall part of assessing student learning, accrediting agencies still look to see that institutions are doing something to evaluate the quality of their faculty members. Likely just as recurring are faculty concerns about and arguments against student ratings: that students lack the qualifications to judge

teaching competence, that student ratings lack reliability and validity, that ratings are highly correlated to the students' grades, that extraneous factors and conditions affect the ratings, and that student ratings are simply popularity contests, among others (Cohen, 1986; Feldman, 1997). In fact, a new website called "ratemyprofessor.com" has just appeared on the internet with a rating system that reinforces the fear that student ratings are really little more than a popularity contest.

Dowell and Neal (1982) characterized the literature on students ratings of teaching as "(1) very extensive, (2) contradictory, and (3) of highly variable quality" (p. 51). What was true over 20 years ago is just as true today. Clearly, the issues related to student ratings have yet to be resolved. So just how useful are student ratings of faculty and courses for both formative and summative evaluation, and how much weight should faculty and institutions afford them in the larger framework of the process of improving teaching and learning or in terms of promotion and tenure? Does the posting of the ratings in student publications or on the internet transform the process into merely a popularity contest, or should such practices continue? These are just some of the questions that underlie the research that has continued to occur in the area of student rating of teaching (SRT).

SRT research generally falls into three broad concerns focused primarily on the rating instrument: reliability, generalizability, and validity. Reliability concerns the extent to which a survey instrument is consistent from one administration to the next. For student ratings, researchers usually use item analysis (internal consistency reliability) and studies of interrater agreement at one time (consistency) and over time (stability) to determine item reliability (Aubrecht, 1981; El-Hassan, 1995). Generalizability relates to the confidence level that the data reflect the instructor's teaching effectiveness across courses rather than just one

particular course in one particular term (El-Hassan, 1995). The most unsettled area involves validity. Greenwald (1997) classifies four types of validity concerns: consequential validity (whether institutions/faculty use ratings to benefit the educational system), conceptual structure (whether ratings are conceptually unidimensional versus multidimensional), convergent validity (how well the ratings measures correlate with other indicators of effective teaching), and discriminant validity (whether variables unrelated to effective teaching influence ratings).

Within the category of discriminant validity, researchers have considered a number of factors as potential biases (variables unrelated to effective teaching that influence ratings). Dividing these biases according to whether they are class, instructor, or student related provides a rough classification of these biases although some biases, such as grading leniency/grade expectations, involve an interaction between instructors and students. When it comes to class-related factors, only the time of day for the course consistently appears in the literature as being a non-biasing factor. The class-related factors that could represent potential biases include level of the course, class size, workload/difficulty, administration and purposes of the ratings, and academic discipline, which represent potential biases. Studies involving the student-related factors of personality, age, student gender, grade point average, and student learning styles have found them to be non-biases. However, factors that at least some studies consider biases include prior subject interest, reason for taking the course, level of the student, instructor gender, student motivation, consumerism, feelings of indebtedness, retribution, fear of retaliation, instructor reputation, and halo effect. Among the instructor-related factors that appear in the literature, research has ruled some out as biases, including age, race, and research productivity. Other instructor-related factors, however, represent

potential biases: rank, years of experience, type of instruction, grading leniency/grade expectations, personality, enthusiasm/"Dr. Fox" effect (instructors receiving high ratings based on their presentation skills rather than based on the content), and likeability (rapport). All of these studies are quantitative as is most of the SRT literature. Additionally, most studies concern survey instruments rather than student comments, and the few studies that examine student comments do so in respect to ascertaining how closely student comments reflect responses to objective questionnaire items.

Most research literature focuses primarily on the reliability, validity, relationships to other variables, and potentially biasing factors related to the rating instruments. Yet recent studies by Nasser and Fresko (2002) and Spencer and Schmelkin (2002) indicate, only now is empirical research beginning to appear that studies administrators, faculty, or students. Feldman (1982) asserted that "more needs to be known about what goes on in students' minds and what aspects of teachers and courses are salient to them as they arrive at assessments of their teachers" (p. 57). According to Wigington et al. (1989), little research to determine what variables influence a student as he or she evaluates the instructor has been conducted. Based on his study of implied performance theories of teaching and their effects on the validity of student ratings, Kishor (1995) calls for focusing on how the students actually form their ratings to improve student ratings, asserting that researchers should apply more effort to students' rating strategies rather than on instrument development. "Understanding how students cognitively form their ratings will most likely provide additional ways of enhancing the validity of students' ratings of instructors" (p. 193). While Williams and Ceci (1997) assert that "we psychologists know very little as yet about *how* students arrive at their judgments about teaching effectiveness" (p. 22), McKeachie (1997)

calls for qualitative research “on what goes on in students’ minds when they are filling out evaluations” (p. 1223). Delucchi (2000) echoes this sentiment, contending that “we need qualitative data on criteria students use when completing course evaluations” (p. 229). The apparent need for qualitative studies accounting for what students think as they complete faculty and course ratings has led to the study at hand.

Furthermore, in an article discussing the influence of grading leniency on SRTs, Greenwald and Gillmore (1997a) assert that “additional research can profitably seek other inappropriate influences on ratings” (p. 1209). In particular, an influence about which research appears limited at best arises that may partially justify faculty concerns: student integrity. This is not to mention that no studies appear questioning institutional integrity concerning the reasons behind asking students to rate instructors and courses as well as concerning the uses of the results. Phillips (1998) in a qualitative study of student evaluations of faculty asserted that another area for further exploration “is the relationship between students’ conversations about teaching effectiveness and the practice they use to quantify their responses” (p. 113). In her suggestions for further research, Phillips commented that students interviewed revealed that they gave instructors high evaluations because they did not want to hurt their feelings. Although she indicated only that these results suggest a need for a different method of evaluating instruction other than quantitative surveys (p. 114), these findings would also warrant further investigation about student candor and sincerity. Does the desire to avoid hurting an instructor’s feelings extend to being less than candid in written comments about the instructor on student rating forms? What if some students are insincere by intentionally either under- or overrating instructors (i.e., they give a lower rating than they believe is warranted to “punish” or “repay” the instructor for low

grades, or they give a higher rating for all factors based on a high rating for one factor due, known as a “halo effect,” or because they feel sympathy for the instructor or because they want to reward an instructor for grading leniency or popularity)? Are Phillips’s findings isolated occurrences, or is the desire to avoid hurting an instructor’s feelings a bias that faculty and institutions must consider when assessing students’ perceptions of instruction? The purpose of this study will be to determine what factors, if any, affect student integrity (candor and sincerity) in completing ratings of teaching and, more significantly, if students are less than candid or sincere in completing ratings, why it occurs as well as what faculty and administration could do to promote these values.

Statement of the Research Problem

In Fall 2001, I conducted a pilot survey of 30 community college students. The study aimed at trying to determine whether any factors affect student sincerity in rating faculty. The results indicated that, indeed, some students are not candid for various reasons, including concerns about anonymity and/or retaliation from the instructor, and that some students overrate instructors to spare their feelings. If students choose to withhold their true feelings or to be insincere, these actions may negate the validity of the results, assuming that they are valid to begin with. Despite the reliability and apparent validity of a given instrument, the results could become greatly skewed (either positively or negatively) if enough students evaluating the same instructor in the same class choose to withhold their true views and/or intentionally under- or overrate the instructor. More importantly, if some students are less than candid or are insincere, then using student ratings as a factor in the promotion and tenure process or as part of a merit pay system may need reexamination. Student ratings of faculty and courses constitute a valuable component of the assessment process—provided that the

results are trustworthy. If the results are not trustworthy, the solution, however, is not to discard the system as some faculty members advocate. Rather, the solution is to seek an understanding of what factors affect candor and sincerity in the first place and to work to create an environment within both the classroom and the institution that promotes student integrity when they are participating in the process.

Therefore, the problem in this study involved identifying the factors that affect student responses when completing student ratings, particularly factors that impact integrity. Through interviewing students about these factors—what causes them and what instructors and the administration might do to promote candor and sincerity, i.e., trustworthiness, in student ratings of faculty—a theory emerged related to the abuse and use of power and the impact on the classroom climate. Clearly, SRTs will be a fixture of the evaluation process for the foreseeable future. Faculty members and administrators should have a vested interest in promoting valid results and therefore should become aware of these factors affecting student integrity and what they might do to promote it.

Research Questions

Participants in this study considered the following questions:

- Are students totally candid when completing student ratings of teaching, or do they over- or underrated an instructor?
- If students choose to withhold their true comments and/or choose to over- or underrate instructors, why have they done so?
- What circumstances or conditions cause students to feel they cannot be candid or sincere in completing SRTs?

- What could instructors and/or administrators do to encourage students to be more candid and sincere in their ratings?

Definition of Terms

Candid: Being candid or having candor refers to the act of being frank or forthright in expressing one's opinions and/or observations; not withholding what one thinks when writing comments on a rating form.

Sincere: Being honest, truthful, and trustworthy; not intentionally changing or distorting one's views on a rating.

Integrity: For the sake of economy in this study, having integrity will mean that one is being candid and/or sincere; conversely, to lack integrity will mean withholding comments or to be consciously or intentionally insincere in completing a rating. In other words, integrity is an interchangeable term for candor and/or sincerity.

Student Rating(s) of Teachers (SRT/SRTs): The literature refers to faculty/course ratings using various terms. Cohen (1997) uses the term "academic evaluation." Commonly, researchers refer to the act as "student evaluation of teaching" (Delucchi, 2000; Marsh & Roche, 2000; Simpson, 1995). The other popular term involves "rating(s)": "students' ratings of instruction" (El-Hassan, 1995), "student ratings of teaching (SRT) inventories" (Jackson et al., 1999), and "student ratings of instructors and courses" (Williams & Ceci, 1997). The terms "evaluation" and "rating" appear almost equally, depending on the preference of the researcher. "Evaluation" may presuppose that students have a common norm or set of values that they use when assessing courses and/or instructors but which they may not actually possess. Additionally, some faculty members believe that students lack the expertise to "evaluate" their teaching, and, therefore, "evaluation" evokes a negative

connotation in the context of “student evaluations of teaching.” The terms “rating” or “ratings” remove the negative connotation associated with evaluation as well as the idea that students are applying a common set of values related to instruction. In the sense that “rating” relates to ranking or grading and because some institutions publish the results of student ratings, in which instructors are ranked according to a number or even a letter grade, the term “rating” may be more accurate for what occurs in institutions that do use the ratings to rank, despite, perhaps, carrying its own form of negative connotation. This study will use the term “student rating(s) of teaching” or its acronym (SRT or SRTs) to refer to any attempt to elicit a student evaluation or rating or opinion about a course and/or an instructor. The term “rating” perhaps reflects more accurately what occurs and it is a shorter word, thus more economical. These ratings may come from any instrument designed for that purpose, ranging from instructor or institutionally generated to nationally normed instruments.

Significance of the Study

The issues related to the use of student ratings of teaching continue despite a plethora of studies. Studies deal with the dimensionality, validity, reliability, and generalizability of students’ ratings as well as with the potentially biasing factors that may affect ratings validity. The number of recent sources attests to the fact that these issues have yet to find a resolution. Furthermore, the continued efforts of researchers to explore the topic of student ratings of teaching speak to the importance of this area of study.

Because SRTs continue to play an important role in providing one measure of teaching effectiveness as well as serve at some institutions as a factor in tenure, promotion, and/or merit pay, resolving concerns about the worth of continuing this practice warrant continued investigation. Although a number of studies have already investigated possible biasing

factors related to students, Delucchi (2000) calls for qualitative data on criteria students use when completing course ratings. Despite the variety of issues related to student bias, a gap appears to exist concerning the issue of student integrity in completing these ratings. Therefore, this study investigates whether students are candid and sincere, ascertains factors that relate to integrity, and provides reasons students choose not to be open and forthright as well as what faculty members might do to promote student candor and sincerity as students complete SRTs. Based on the results, this study puts forth a grounded theory concerning the abuse and use of power and how that influences discourages or promotes student integrity.

Researcher's Perspective

Having served as a faculty member in community colleges for over 20 years, I experienced student ratings of my classes and of me through a variety of instruments, both institutionally developed and nationally recognized instruments. Many of the faculty concerns about the reliability and validity of student ratings found in the literature are ones that I and/or my colleagues have shared. In my position as dean of instruction at Lamar Community College, I found that faculty members continued to raise concerns over the reliability and validity of student ratings. Over the years, students have also reported to me that they have either withheld what they truly thought about a particular instructor or gave instructors higher ratings than they deserved for various reasons.

I strongly believe that assessment of faculty members is a critical component of a continuing effort to improve teaching and learning. A variety of means exist for gathering students' perceptions of instructors and what they think they are learning, including various Classroom Assessment Techniques as well as formative and summative student ratings of teaching. Although some institutions use the results of SRTs for promotion, tenure, and pay

purposes, my perspective is that institutions should approach using these results with caution, a perspective that this study reinforces. Faculty members and institutions should work to develop classroom environments and institutional cultures that encourage students to be candid and sincere in providing their perceptions in completing SRTs. To accomplish this goal will require more information about what students are thinking when they complete SRTs.

Additionally, my perspective, based on personal reports from students and on the pilot survey I conducted in Fall 2001, is that some students choose to withhold their true thoughts and/or intentionally over- or underrate some instructors when completing SRTs. My initial question was whether this occurrence is relatively isolated or pervasive enough to affect the validity of SRTs. Also, if the practice is widespread, trying to understand what students are thinking when they withhold their comments and/or over- or underrate instructors may help faculty members and administrators improve this important source of information for the improvement of teaching and learning. Based on my findings, a lack of candor and sincerity is not only pervasive but also varies in terms of the students' relationship with the instructors as a given student may be totally candid and sincere with one instructor and withhold comments with another. Instructors and, to an extent, administrators create either a climate of trust or a climate of fear as the result of various actions and characteristics. Additionally, fear manifests itself in a variety of ways. However, the likelihood of a lack of integrity is far greater in a climate of fear than in a climate of trust. From this investigation a grounded theory concerning the abuse and the use of power emerged that accounts for which climate an instructor creates in the classroom. Through recognizing this dynamic that affects the interaction between instructors and students, both instructors and administrators can take

steps to create an atmosphere that promotes candor and sincerity so that student ratings of teaching are meaningful and fully contribute to the improvement of teaching and learning in higher education.

CHAPTER TWO

REVIEW OF THE LITERATURE

The subject of student rating of teaching (SRT) has been widely studied from so many perspectives that one might wonder why the issues have yet to be resolved. Ory (1990) cites Cashin (1988) who had asserted that over 1,300 articles and books dealing with research on student ratings already existed. Ten years after Cashin, Wilson (1998) claimed that the number had swelled to over 2,000 studies, “making it [student evaluations] the most extensive area of research on higher education” (p. A12). Researchers continue to investigate a variety of issues ranging from the debate over the multidimensional versus global nature of the questions in the instruments to reliability and generalizability to validity factors that may bias the results. This interest is understandable given continued mistrust of the results by some faculty and given the continued use of and emphasis on SRTs to provide diagnostic information to faculty about their teaching, to measure faculty effectiveness for tenure/promotion/merit pay decisions, to inform students about courses and instructors as they make their class selections, and to provide “an outcome on a process description for research or teaching” (El-Hassan, 1995, p. 411). The interest is also important when faculty raise questions such as the one Greenwald and Gillmore (1998) present, a question that may represent the central question in the minds of many in higher education concerning the use of student ratings of teaching: “Would higher education work better in the absence of student ratings?” (p. 1229). As they point out, arriving at an empirical answer to this question is

“very difficult.” Hence, the number of studies in search of answers is likely to continue to grow in the foreseeable future, keeping pace with the amount of skepticism that still exists among some faculty members.

Trout (1997) claims that professors can easily raise evaluation scores merely by complying with student demands, which he asserts are calls for “dumbing down their courses” (p. 30), and Martinson (2000) echoes this sentiment when he comments on the dysfunctional aspects of relying on student evaluations concurrently with pressure to increase academic standards, implying that increasing standards will create an adverse reaction among students resulting in lower student ratings. In response to and support of Trout’s comments, Bauer (1997) remarks that SRTs are not valid measures of teaching effectiveness and violate academic freedom. “Many professors suspect that their student ratings fluctuate systematically in accordance with factors that have little to do with the actual content of their lectures, reading materials, or grading policies” (Williams & Ceci, 1997, p. 14). Cohen (1997) expresses a variety of concerns about the process of student ratings of instructors, including students being unqualified to judge an instructor’s mastery of the subject matter and students retaliating against instructors for refusing to consider grade changes or for not being available for students’ questions when it is convenient for the students. Moreover, Cohen even raises the question about whether anonymous student ratings trample professors’ constitutional rights while protecting students from possible “professorial reprisals.” However, recent research by Nasser and Fresko (2002) studying faculty views of student ratings of teaching find that instructors receiving higher ratings tend to view SRTs as more reliable and valid than those receiving lower ratings. These results lead them to conclude that instructor opposition may be not only ideological but also a self-interested rationalization.

“Instructors support course evaluation when it flatters them and oppose it when they are criticised” (p. 197).

In trying to take a more objective approach to determining faculty attitudes toward SRTs, Marsh (1987) reports on a study he conducted in which he surveyed faculty to determine which of a list of 17 characteristics they believed would cause bias in student ratings. The following percentages occurred for these factors: course difficulty (72%); grading leniency (68%); instructor popularity (63%); prior student interest in subject (62%); course workload (60%), class size (60%), reason for taking the course (55%); and student’s GPA (53%) (p. 305). Such reactions reflect what some faculty members either voice openly or think secretly about student ratings of teaching—that they lack reliability and/or validity.

Although educators have used student ratings of professors with the goal of improving teaching since the early 1900s (Cruse, 1987), the concerns with the reliability and validity of both the instruments and the results emerged in the 1970s, peaking in the early 1980s (Greenwald, 1997). Ory (1990) categorizes the 1960s as the period in which students’ demands for faculty accountability and course improvement first changed the administration and use of student ratings, whose purpose had been mainly to help students select professors and courses prior to that decade. Historically, according to Greenwald (1997) the major research concern of the 1970s was the possible effect grades have on ratings as the belief was that grading leniency resulted in higher student ratings, i.e., a grade-induced bias was occurring. Greenwald asserts that the experimental tests that have occurred support the hypothesis that grading leniency-strictness affects ratings and that despite the questioning of these conclusions by critics, no studies have empirically refuted those conclusions. Ory

(1990) characterizes the 1970s as a period in which administrators began a more systematic collection of student ratings to use in decision-making processes. The juxtaposition of purposes here is an interesting one. A decade marked by an increased administrative use of ratings corresponds to a period in which research studies focused on seeking to identify biases such as grading leniency. It appears clear that the use of ratings may have had a profound effect on research concerns. Most of the research studies related to SRTs have been correlational construct-validity designs involving three kinds of studies: multisection validity studies, path-analytic studies, and multitrait-multimethod studies. According to Greenwald (1997), "In summary of the relatively recent literature on student ratings . . . prominent reviews published since about 1980 give a clear impression that major questions of the 1970s about ratings validity were effectively answered and largely put to rest by subsequent research" (p. 1184). Ory (1990) contends that one could label the 1980s as the peak period for administrative use of student ratings, with 1990s seeing most educational institutions using student ratings at least in part in regard to merit and promotion. Despite the general conclusions from the research as a whole that student ratings are reliable and valid, controversies and unanswered questions persist (d'Apollonia & Abrami, 1997). It is only recently that the focus of any research is shifting from concerns about the reliability and validity of the instrument to the focusing on the issue of validity from a cognitive information processing perspective with a focus on the students rather than on the instrumentation (Kishor, 1995).

Kishor cites Anderson (1981) who asserts that cognitive processes of attention, storage, retrieval, and integration are all involved in producing numerical ratings as required on the Likert-type instruments used in almost all student rating instruments. This means that

students' rating judgments may be valid only to the extent that a rater follows rational procedures in forming those judgments. Consequently, the validity of the ratings will depend on students' ability to make inferences when they complete the performance rating instruments, and these inferences may contain bias—a bias in judgment. According to Kishor's approach, the raters themselves can be the source of bias. If biases in performance ratings result from the operation of implicit personality theories that influence the cognitive processing of performance information, then it is unlikely that even psychometrically sound rating instruments will provide valid evaluations. Hence, the need arises to focus attention on raters' cognition or how they cognitively arrive at their ratings. In other words, it is important to understand student thought processes and what influences their ratings, for as Kishor indicates, it is perfectly conceivable to have a reliable and valid instrument that still yields untrustworthy results due to invalid student rating judgments. Despite such findings, most research continues to deal with the ratings instruments.

Although reliability and generalizability studies appear in the literature, most of the studies relate to one of four validity issues: consequential validity (whether institutions/faculty use ratings to benefit the educational system), conceptual structure (whether ratings are conceptually unidimensional versus multidimensional), convergent validity (how well the ratings measures correlate with other indicators of effective teaching), and discriminant validity (whether variables unrelated to effective teaching influence ratings) (Greenwald, 1997). What follows in this chapter is a review of some of findings appearing in the literature related to these four general areas. These areas will provide an overview of major concerns related to SRTs; however, the aspect of discriminant validity relates most closely to factors that affect student integrity in completing SRTs, requiring a more extensive review of

some of the research that has appeared.

Studies Concerning Consequential Validity

In relation to the other validity concerns related to student ratings of teachers, the one least discussed is consequential validity. The issue here relates to how effectively ratings are put to use (Greenwald, 1997). As McKeachie (1997) indicates, most people concerned with improving teaching consider the responses that students provide through student ratings related to specific teaching factors as helpful for that improvement. However, he also contends that “[g]eneral overall ratings provide little guidance” in using student ratings for the purpose of promotion and tenure (McKeachie, 1997, p. 1218). Part of the problem concerns which approach to ratings—multidimensional evaluations of factors or global evaluations of general teaching skill (discussed below)—provides the most useful information, and some, like Abrami (1985), recommend against the widespread use of SRTs for promotion and tenure decisions. McKeachie (1997) advocates having students rate their attainment of educational goals. In this area of study, student ratings of teaching become one part of comprehensive faculty assessment systems. For example, Braskamp and Ory (1994) discuss a variety of means of collecting trustworthy evidence to assess faculty, including not only student ratings but also written appraisals; rating scales and checklists; interviews; observations and videotaping; indicators of eminence, quality, and impact; achievement and outcome measures; and records and portfolios.

As McKeachie (1997) sees it, the major validity problem resides in the use of student ratings by personnel committees and administrators. He contends that these committees and administrators fail to give SRTs “heavy weight” in promotion decisions (p. 1222).

Additionally, his concerns extend to institutions' attempting to compare instructors with each other by using numerical means or medians. Feldman (1997) adds that promotion and tenure committees may want to take into account that it may be more difficult to teach effectively in some courses rather than others.

In a much earlier review of other studies, Kulik and McKeachie (1975) present three general research conclusions that administrators should consider before putting too much emphasis on student ratings for the purposes of tenure and promotion: ratings reflect more than teaching skill (i.e., various biases influence ratings), ratings are only partial indicators of teaching effectiveness (i.e., other factors such as noncourse teaching, research supervision, direction of independent studies, or advising should count for administrative decisions), and ratings for the purpose of tenure and promotion may impact the general academic community atmosphere (i.e., use of ratings may create an atmosphere of distrust, hostility, resentment, etc.).

Applying generalizability theory to student ratings, Gillmore et al. (1978) recommend averaging ratings across at least five courses to enhance generalizability so that administrators base any promotion and tenure decisions on a composite of an instructor's career. Abrami (1989a) in his discussion of the global versus multidimensional issue, discussed in more detail below, raises the concern that when administrators do use student ratings, we cannot expect them to weigh properly the information that factor scores provide in deciding about the quality of an instructor's teaching.

The research concerns for this validity issue involve studying what teachers can do to enable students to become more sophisticated raters. Furthermore, training needs to extend to faculty on how to educate their students on these issues as well as to encourage faculty to

explain to their students why what they are doing is likely to result in better learning. Finally, a need for research exists on how to train personnel committee members and administrators to become better evaluators as well as on ways of communicating the results of student ratings to improve the quality of their use (McKeachie, 1997).

Baker and Copp (1997) add to McKeachie's contentions by asserting that other factors influence students' perceptions and that promotion and tenure committees should consider students' subjectivity. "By failing to consider students' subjectivity as such, colleagues may promote gender bias in the university" (p. 41). Committees should be careful about taking students' comments at face value and should consider students' gendered expectations of female and male professors. They should also consider the impact of classroom conditions on professors' ability to fulfill students' expectations. Standard course evaluations omit the social context of the class, thereby excluding the meanings that are attached to social interaction.

Certainly, the whole issue of using student ratings for the purposes of promotion and tenure has provided an impetus for the vast number of studies related to proving the validity of student ratings. Even as long ago as 30 years, Scott (1975) asserted the importance of providing validity studies that would demonstrate a positive correlation between information from student ratings and student achievement "if widespread use of student ratings for merit, promotion, and/or instructional improvement is to be continued" (p. 445). Others, such as Cohen (1983), have cautioned against student ratings being used as the sole criterion of effective teaching for promotion and tenure decisions.

Ultimately, the concerns with consequential validity pivot around arriving at rating results that are meaningful to faculty and that personnel committees and administrators use

with proper emphasis, not only to promote and tenure faculty or to provide merit pay but also to improve teaching and learning. Research related to the uses of student ratings for the purpose of improving instruction, as reported in a meta-analysis that Cohen (1980) conducted, indicates that student feedback, particularly mid-semester ratings, do have a positive impact on improving instruction. Perhaps if student ratings were used only for formative ends—for the purpose of improving teaching and learning—concerns with the consequential validity would be fewer. However, as long as institutions continue to use ratings as part of summative evaluation, this issue of consequential validity will persist. To achieve the end of arriving at meaningful results for all purposes requires that those involved trust the results of student rating instruments, which involves the second validity concern: the conceptual structure.

Studies Concerning Conceptual Structure

The debate over the conceptual structure of ratings is not over the nature of teaching. Researchers agree that teaching is a complex, multidimensional activity (Abrami & d'Apollonia, 1991; Burdsal & Bardo, 1986; Cashin & Downey, 1992; Feldman, 1997; Marsh & Roche, 1993). Furthermore, findings show that the instructor, rather than the course taught, is the primary determinant of student ratings (Gillmore, Kane, & Naccarato, 1978; Marsh & Overall, 1981). The debate hinges on the issue of the validity of multidimensional versus the global nature of the questions and intent of the instrument. This debate arises in relation to the use of SRTs. As mentioned earlier, SRTs serve the general purposes of providing instructors with information about their teaching, providing information for students to use in selecting courses and instructors, providing information for research on teaching, and providing a summative measure for personnel decisions. According to Marsh

and Hocevar (1991), general agreement exists about the value of using multiple dimensions for the first three purposes. The controversy over employing multidimensional versus global questions arises in relation to the fourth purpose—using SRTs for summative evaluations in personnel decisions.

Those who support multidimensionality contend that ratings are conceptually and empirically multidimensional, just like effective teaching. In other words, multidimensional survey instruments contain several items under multiple dimensions of instruction. For example, the Students' Evaluation of Educational Quality (SEEQ) instrument contains nine dimensions or factors: Learning/Value, Instructor Enthusiasm, Organization/Clarity, Group Interaction, Individual Rapport, Breadth of Coverage, Examinations/Grading, Assignments/Readings, and Workload/Difficulty (Marsh & Hocevar, 1991; Marsh & Roche, 1997). Each of these factor areas contains several questions to assess these dimensions. A faculty member might receive a high rating in Instructor Enthusiasm, for instance, but low in Organization/Clarity, thus enabling the faculty member and the administrator to discriminate an instructor's strengths and areas for improvement according to these factors. Of course, one apparent drawback with multidimensional SRTs is that they are longer, thus requiring more time to score and interpret. Additionally, their length may cause some students to hurry through the ratings, thus making the results more susceptible to halo effect (the tendency of a rater to allow an overall impression to affect similar ratings for all dimensions). (A more extensive discussion of halo effect appears later.) However, proponents, such as Marsh and associates, argue a multidimensional approach provides a better analysis of the strengths of a particular instructor as well as areas to improve. A global ratings instrument contains few questions and generally asks students for their overall rating of an instructor or a course

without asking about specific dimensions. For instance, a global question might read something like this: How would you rate this instructor in overall ability? Often, the multidimensional instruments contain a few global questions as well. Those supporting global ratings argue that a single score can best represent students' ratings (Marsh, 1987).

In their study of the student perceptions of teaching effectiveness (SPTE), developed at Wichita State University, Burdsal and Bardo (1986) discovered that the use of three single-item global indicators of teaching effectiveness (overall course evaluation, general ranking of the instructor, and the degree to which a student would recommend the instructor to other students) resulted in significant mutual intercorrelation but all having their highest loadings on different primary factors. They conclude that "such single items are overly simplistic representations of students' perceptions of the instructor" (p. 75). Furthermore, they advise against the use of single item indicators for the use of tenure, promotion, or salary considerations because using single-item indicators is "risky and questionable." Aleamoni (1987) goes as far as asserting that believing that student ratings on single general items are accurate measures of instructional effectiveness is a myth.

A recent study by Jackson et al. (1999) supports the validity of a multidimensional approach and concludes that common dimensions exist along which students evaluate their courses and that these factors provide a "comprehensive basis for evaluating students' perceptions of instruction" (p. 594). El-Hassan (1995) conducted a multisection validity study and found that the results favor the use of specific dimensions over the use of global items as they have higher correlations with achievement criteria. Marsh and various colleagues have conducted the most extensive research in this area, developing the SEEQ

instrument. Numerous studies support the multidimensionality of SRTs (Marsh, 1982a, 1982b, 1983, 1984, 1987; Marsh & Dunkin, 1992; Marsh & Hocevar, 1984, 1991). Feldman (1988), in his meta-analysis comparing student and faculty members' perceptions of effective teaching characteristics, asserts that correlational analysis between specific and overall SRTs provides "prima facie evidence of the importance of various specific instructional characteristics in discriminating among students' global assessment of teachers" (p. 315). Marsh and Roche (1997) contend that, "Global or overall ratings cannot adequately represent the multidimensionality of teaching" (p. 1188) and that global ratings may be more susceptible to context, mood, and other potential biasing factors. In other words, because the multidimensional items relate more closely to actual teaching behaviors, potential biasing factors are less likely to influence them.

Opponents of a multidimensional approach, such as Abrami and d'Apollonia (1991) and Feldman (1976b), claim methodological problems exist with using factor analysis as evidence for the construct validity of distinct instructional dimensions, and they raise equally compelling arguments in favor of using global ratings. According to Abrami (1985), the salient question should not be what the ideal college teacher is but rather what the ideal college teacher is for different contexts and for different goals, objectives, or desired outcomes of instruction. He contends that some multidimensional items, such as ones that related to promoting student discussion or interaction may be inappropriate in certain circumstances, for example, with large or with lecture classes as opposed to small classes and discussion classes, thus providing one more reason for advocating the use of global items for rating instructors.

In later work, Abrami (1989a; 1989b) articulates five concerns with the reporting and

use of separate factor scores for summative purposes: (1) We lack sufficient evidence to establish either what the dimensions of effective teaching are or whether and how they interrelate; (2) problems may exist concerning the content validity of specific items and some of the dimensions they compromise when ratings are used across a wide variety of courses, instructors, students, and settings, e.g., using rapport and interaction in large classes than small, discussion or studio classes than lecture, etc.; (3) the construct validity of some rating factors is poor; (4) we know less about the generalizability of the specific rating factors than global ratings; and (5) we cannot expect administrators or nonexperts in evaluation to weigh properly the information factor scores provide in arriving at a single decision about an instructor's teaching quality. His concerns lead him to favor global rating items or a carefully weighted average of rating factors in place of separate factor scores for summative evaluations of faculty (Abrami, 1989b). The debate, however, does not end here. Marsh (1991) responds to the criticisms of Abrami and d'Apollonia (1991) arguing that they fail to operationalize criteria for unidimensionality or multidimensionality and that they misunderstand Marsh's use of hierarchical confirmatory factor analysis. Marsh is also critical of their criticism of the generalizability of his findings, the appropriate use of SRTs, multisection validity studies and manipulations of instructional characteristics, and Abrami and d'Apollonia's (1991) secondary analysis of students' evaluations of educational quality data.

Performing their own principle-components factor analysis using the instructor's role in delivering information, the instructor's role in facilitating a social learning environment, the instructor's role in evaluating student learning, and four miscellaneous teaching behaviors (maintaining discipline, choosing appropriate required materials, knowing the content, and

using instructional objectives) along with a second-order factor analysis, Abrami, d'Apollonia, and Rosenfield (1996) found that student ratings of instruction measure General Instruction Skill. General Instructional Skill consists of the three subskills of delivering instruction, facilitating interactions, and evaluating student learning. According to d'Apollonia and Abrami (1997), an equally viable interpretation of factor studies, such as Marsh's, is that student rating forms measure a global component—General Instructional Skill.

Those researchers favoring a global dimension (Abrami, 1985; Abrami, d'Apollonia, & Rosenfield, 1996; Cashin & Downey, 1992; Cohen, 1986) advocate using global items, such as questions about the overall effectiveness of the instructor or a carefully weighted average of rating factors since no general agreement on the dimensions of effective instruction exists across all student rating forms. As Abrami (1985) argues, administrators or tenure committees making summative evaluations of instructors do so on a unidimensional basis rather than multidimensional and that even with multidimensional results that show variations in factors such as organization, concern, knowledge, and so forth, the evaluators still assemble these results into a single index of teaching effectiveness. Cashin and Downey (1992) found that two global items (one concerning the instructor's being an excellent teacher and one concerning the course) accounted for more than 50% of the variance in the weighted composite criterion measure. Their results led them to conclude that institutions can use single global items for summative purposes, reserving the multidimensional items for the purpose of teaching improvement.

Besides the issue of multidimensionality versus global approaches, Jackson et al. (1999) point out the existence of two major issues lacking consensus regarding the construction of

student ratings of teaching: the item content and the number of items. Researchers can design student rating instruments for either a narrow range of applicability, such as to evaluate a music performance, clinical classes, etc., or to evaluate dimensions common to a broad spectrum of subject matter and course formats. The purpose of gathering student comments—whether to provide formative information to improve teaching, summative information to promote or tenure, or both—can also affect questionnaire design. Additionally, some of the lack of consensus occurs due to variation in factor analytic procedures as well as to differences in identifying the dimensions of student perceptions of teaching effectiveness.

A recent study by Palmer (1998) finds that attribute categories similar to those Marsh uses in the SEEQ are not independent, with two constructs (interaction and organization) accounting for 78 percent of the variance. Palmer also discovers a strong correlation between the indicators of general student satisfaction with the quality of instruction and these two constructs. These results led her to conclude that if the purpose of student ratings is formative, i.e., to improve teaching, a multidimensional instrument is more suited to provide diagnostic information whereas if the purpose is summative, i.e., to provide an basis for evaluating an instructor for tenure and promotion, then administrators could use global items because her study finds that student responses to global items correlate highly to attribute items.

An early study by Doyle and Whitely (1974) also supports these conclusions. They find that ratings of overall course and instructor “effectiveness” vary significantly across certain student groupings and therefore are not generalizable. “The more abstract or all-encompassing the item, it seems, the more likely it is to be influenced by individual

differences among students, but the averaging of ratings appears to control these ‘biases’” (p. 271). If evaluations are for personnel decisions, they advocate using the mean with student disagreement being viewed as error, but if used for diagnosis of teaching, then student differences provide important information, and individual items are the proper unit of analysis.

Ultimately, the deciding factor as to whether to use multivariate (or multidimensional) or global ratings revolves around the issues of the purpose for using student ratings and whether one approach is more reliable and valid than another. The debate between those who advocate each position appears still unresolved and perhaps will never be resolved as long as institutions continue to use student ratings both for improving teaching and for job retention, tenure, and promotion. However, those who support a multidimensional approach must also address the third and fourth major concerns: convergent and discriminant validity.

Studies Concerning Convergent Validity

How well ratings measures correlate with other indicators of effective teaching constitutes the focus of convergent validity, which involves the construct validation of students’ ratings. Freedman, Stumpf, and Aguanno (1979), in their validity study of the Course-Faculty Instrument (CFI), distinguished between “intrinsic” and “extrinsic” variables. Factors that instructors can control, students can directly observe, and relate both theoretically and empirically to student perceptions of their learning constitute the intrinsic variables. In other words, controllable factors that appear directly related to student learning are intrinsic and serve as convergent criteria. The extrinsic variables consist of variables beyond instructor control, such as gender, experience, rank, class size, etc., and any factors students cannot observe as well as factors such as publishing activities that differentiate high

and low learning experiences in terms of student perceptions but that students may not realize as they take a course from a given instructor. These extrinsic variables fall into the area of discriminant validity—or biases—which is the topic of discussion in the next section.

Since the 1980s, most of the research on student ratings has concerned this search for convergent validity (Greenwald, 1997). However, this research has had its critics. For example, Dowell and Neal (1982) assert that despite the large number of studies, few are interpretable because of methodological issues related to section means, controls for student ability, and criterion inadequacy, and Aleamoni (1981) indicates that opinions about SRTs range from their being valid, reliable, and useful to invalid, unreliable, and useless. Four reasons for such a range of opinions are that those who study student ratings have preconceived biases, that some have unrealistic expectations about what student ratings can and should be able to do, that a plethora of *ad hoc* instruments based on varied item content and untested psychometric properties exist, and that both student-evaluation instruments and the research based on them are the result of a fragmentary approach (Marsh, 1987).

Generally, studies have shown student ratings to have moderate convergent validity (Cohen, 1983; Greenwald & Gillmore, 1997). As Dowell and Neal (1982) characterize the situation, the literature fails to support a position that ratings are sufficiently valid for policy uses or that ratings are clearly invalid across all validity studies either. However, in response to Dowell and Neal, Cohen (1983) concluded that based on his own research synthesis, student ratings of instruction provide a modest index of instructional effectiveness with students being able to distinguish among instructors based on how much they have learned. Three types of studies generally characterize the correlational construct-validity designs used to prove convergent validity: multisection validity studies, multitrait-multimethod studies, and

path-analytic studies.

Multisection validity studies, such as the one Frey, Leonard, and Beatty (1975) conducted, involve comparing student performance, usually on a common final examination, of students from multiple sections of the same course taught by different instructors. The intent of multisection validity studies is to compare teachers in terms of operationally defined student learning. In the case of the study by Frey et al., for example, they consider seven aspects of teaching (clarity of presentations, workload, personal attention, class discussion, organization-planning, grading, and student accomplishment) and how these factors correlate with performance on the final examinations with their results demonstrating that clarity of presentations, organization-planning, and student accomplishment significantly correlate with performance on the final exam. However, as Marsh (1987) points out, most university teaching does not occur in such a setting. Most studies have tried to match student ability approximately across sections. The basic intent of these studies concerns determining whether the ratings of instructors reflect differences in achievement (as measured by final examination results) for students of different instructors. These studies support an interpretation of observed correlations between grades and ratings in terms of parallel effects of a third variable (teaching effectiveness) on both measures. A correlation of grades and ratings only or mainly, because good teachers produce both high grades and high ratings, supports the validity of student ratings. Generally, multisection studies have produced evidence for modest convergent validity of ratings (Greenwald, 1997).

Freedman and Stumpf (1978) use a multitrait-multimethod strategy to confirm the validity of the Course-Faculty Evaluation Instrument (CFI) that they developed for the Graduate Business School at New York University. A multitrait method, like the one

Freedman and Stumpf used, involves identifying dimensions of instruction. In the case of the development of the CFI, they ultimately identified seven dimensions that they claim are relatively independent of each other: Subject Functionality, Subject Affect, Subject Difficulty, Instructor in Class, Instructor in General, Graded Assignments/Examinations, and Text/Required Readings. Although they find that these dimensions are not independent, they develop these dimensions to correlate highly with perceived learning. For each of these dimensions, they then create 16 or more non-redundant items (96 total). In effect, then, a multitrait strategy aims at validating the various dimensions of teaching effectiveness. Freedman and Stumpf next use factor analysis to validate each dimension according to both the strength of the point-biserial correlations and the internal consistency of the dimensions as indexed by coefficient alpha. Using these methods allows them to eliminate items within a dimension that do not contribute to the psychometric properties of the CFI. The multimethod portion of their study involves using three different question format and scaling methods. The first method, which they conclude is the most reliable, uses a yes (Y)-undecided (?)—no (N) format. Students respond to a quality in relation to a dimension; for example, in the subject difficulty dimension, students would rate the quality “demanding” (yes, undecided, or no). The second method uses short descriptive statements, to which students respond using a five-point Likert scale ranging from “strongly agree” to “undecided” to “strong disagree.” Finally, the third method consists of having students complete a seven-space semantic differential type of scale with bipolar adjectives, such as “Demanding—Undemanding.” Through such a combination of multitrait and multimethod testing, Freedman and Stumpf discover that convergent validity exists in the CFI and that method one produces the best discriminant validity. From their study, they also conclude that student learning perceptions

are “a meaningful and relevant, albeit limited, criterion of educational effectiveness” (p. 199) with students able to provide valid evaluations.

In a following study of the Course-Faculty Instrument (CFI), Freedman, Stumpf, and Aguanno (1979) examine both intrinsic (convergent) and extrinsic (divergent) variables in the process of validating the CFI. They assert that although some correlations of extrinsic variables with the overall instructor ratings, such as instructor gender and class size, may indicate bias, other correlations, such as experience, rank, and publishing activities, may be as suggestive of validity as bias. In general, they conclude that their data support the construct validity of the CFI.

Some multitrait-multimethod studies also attempt to reveal that both convergent and discriminant validity comprise student ratings. In other words, they seek to demonstrate that convergent and discriminant validity correlate relatively well with measures based on other methods for assessment of the construct of quality of instruction and relatively less well with measures assessing other constructs. For instance, Marsh (1982b) uses the same 35-item rating form to elicit ratings from students about their instructors and from those instructors in terms of their own self-evaluations. Since Marsh employs the same instrument for both groups and they show agreement, his study supports the convergent validity of both. Additionally, his results demonstrate the distinctiveness of the different evaluation factors, thus supporting divergent validity. His study provides evidence that student ratings show good agreement with instructor self-evaluations of teaching effectiveness; consistent evidence occurs for the validity of student ratings for both undergraduate and graduate level courses; and the distinctiveness of the different rating scales establishes the importance of using multifactor evaluation instruments that are developed using factor analysis.

A validation study of Hudson's *Teaching Evaluation Form* (TEF) that Pike (1998) conducted takes a slightly different approach than Marsh's in designing a multitrait-multimethod study. The study involves having students in social work programs at two different universities complete the rating forms as well as having the deans at these two institutions also rate the same faculty members. The study uses the deans' ratings to examine the convergent construct validity of the TEF. Studies involving the use of multitrait-multimethods, such as those just discussed, have generally reported evidence for both convergent and discriminant validity of student ratings (Greenwald, 1997).

Path-analytic studies explore the possibility that a third variable other than teaching effectiveness can explain the correlation between grades and ratings. These third variables that various researchers have studied include student motivation (Howard & Maxwell, 1980), prior subject interest (Marsh, 1980, 1983, 1987), expected grade (Marsh, 1980, 1983), higher levels of workload/difficulty (Marsh, 1983), and a higher percentage of students taking a course for general interest (Marsh, 1983). Howard and Maxwell (1980), for example, create three models in their investigation of a causal link between student grades and student ratings: a grading leniency bias model, a teaching effectiveness model, and a student characteristics model. They compute intercorrelations among five target variables (progress, grade, motivation, and two satisfaction items). By running regression analyses using these variables, they examine 10 separate path analyses comparing the student motivation model with the grading leniency model. This process of running different correlation analyses produces potential paths of causation. In Howard and Maxwell's study, they find that motivation rather than grade leniency has a stronger relationship with grades and student satisfaction.

Overall, these various construct-validity studies provide evidence for the validity of factors that appear in the array of student ratings instruments that exist today. Convergent validity exists when different methods of measuring the same characteristic show significant intercorrelations. Multisection, multitrait-multimethod, and path-analytic studies illustrate the three major methods that researchers use in trying to determine convergent validity. Convergent validity represents the attempts to verify the internal factors related to effective teaching. Even more extensive research deals with the issue of exploring the possible effects of external factors that could cause a bias in the ratings. This research falls into the realm of discriminant validity.

Studies Concerning Discriminant Validity

“Perhaps more than any other area of student-evaluation research, the search for potential sources of bias is extensive, confused, contradictory, misinterpreted, and methodologically flawed” (Marsh, 1987, p. 309). That conditions not under the instructor’s control can influence the ratings or that conditions may somehow be “unfair” to the instructor constitute biases, and as Haladyna and Hess (1994) note, researchers need to detect and correct bias in student ratings to prevent misinterpretations and misuses of results. Feldman (1997) defines bias as “one or more factors directly and somehow inappropriately influencing students’ judgments about and evaluation of teachers or courses” (p. 370). In other words, bias involves any condition or influence on students’ attitudes such that evaluations fail to accurately reflect the instruction that students receive. The definition guiding Marsh (1987) in Student Evaluation and Educational Quality (SEEQ) research is that student ratings are biased to the extent that variables unrelated to teaching effectiveness influence them and, perhaps, to the extent that this influence generalizes across rating factors rather than being

specific to a particular factor logically related to the influence. According to Marsh, for a bias to exist, the variable must not only correlate with student ratings and warrant a causal interpretation, but it must also be demonstrated that the variable does not correlate with effective teaching. Using this definition in connection with the “Dr. Fox effect,” for example, Marsh contends that the effect is not a bias because it influences only the enthusiasm factor on SRTs. Additionally, enthusiasm is a factor associated with effective teaching. Studies concerned with discriminant validity examine potentially biasing factors.

Opinions vary about the influences of biasing variables. In a survey of faculty members, Marsh (1982b) reports what instructors believe are the main biasing factors affecting student ratings: course difficulty (72%), grading leniency (68%), instructor popularity (63%), student interest in subject before course (62%), course workload (55%), class size/enrollment (55%), and required versus elective (55%) (p. 266). Marsh and Roche (1997; 2000) believe that potential biases have relatively little affect on ratings and that bias interpretations usually fail to control valid effects on teaching, such as class size and instructor enthusiasm (the Dr. Fox effect), that ratings accurately reflect. D’Apollonia and Abrami (1997) concur, seeing little evidence of bias in ratings. They contend that few characteristics have been shown differentially to affect ratings and teacher-produced student learning. On the other hand, Greenwald and Gillmore (1997a) have found that biasing factors such as grade leniency, small class size, and instructor enthusiasm do affect student ratings. Still another perspective from which to consider bias comes from Dowell and Neal (1982). They point out that attribution theory leads one to expect a variety of situational and student characteristics would bias student ratings. For instance, poorer students who remain in a course until the end of the term may overestimate the amount learned as a means of

reducing the dissonance between their poor performances in relation to their investment of time in the course. Conversely, poor students might attribute their poor performance to external factors such as the instructor, resulting in their underestimating the teacher's ability. However, Gigliotti and Buchtel (1990) found attributional bias as having a minimal to nonexistent effect on evaluations. Clearly, the issue of bias has yet to be settled. Researchers have examined a host of factors that may bias student ratings of teaching. Some of these factors, such as the rank of instructor, appear resolved as not constituting factors that bias ratings while others, such as grade leniency and workload, are still being debated.

A study that Griffin and Pool (1998) conducted may best reveal the paradox of student ratings and the search for validity. Their intent was to determine the impact of conducting more frequent evaluations during a semester on the final student ratings of the course, using classes over a three-and-a-half year period. The experimental courses provided the instructor with comments about how he could improve student learning in his statistics course periodically throughout the semester. The instructor would then summarize the comments and spend the first five minutes of the following class discussing the results and modifications in teaching that would help improve the course. The instructor taught the control courses exactly the same, even incorporating the suggestions of the experimental classes into his teaching of the control classes so that the only difference between the experimental and control classes would be using the *Class Assessment Questionnaires* (CAQ). Results are positive as Griffin and Pool find, in particular, that slowing the lecture rate, providing more examples of concepts discussed, and providing more hands-on activities—suggestions coming from student responses on the CAQ—result in improved student learning as well as in higher end-of-term student evaluations for the course and

instructor. Therefore, based on these results, one could conclude that student feedback throughout the semester can improve student learning, so higher ratings at the end of the course could simply result from better student learning. However, the paradox is that although the control classes learned as effectively and benefited from the teaching modifications that the experimental classes suggested, the end-of-course ratings were not as high as the ones from the experimental classes. One conclusion that the authors report that one could draw is that the end-of-term rating instrument is invalid because non-instructional factors affect the results. Or it could produce the Hawthorne Effect in that the act of giving students the ability to provide comments that the instructor then discusses with them and acts upon actually gives the students a sense of empowerment, which subsequently leads to greater motivation and to higher ratings for the instructor's organizational and planning skills, knowledge of content, course expectations, and skill in evaluating student performance. These results serve to illustrate the dilemma found often in studies related to exploring potential sources of bias in student ratings and which can be summarized in the following question: Do the factors effect improved teaching and learning, in which case they are not biases, or are they, in fact, external and beyond the instructor's control, in which case they are biases?

One could categorize potential biases according to a number of general factors. For example, Haladyna and Hess (1994) designate five categories: student variables (e.g., disposition to instructor, gender, age, course level/year in school, graduate/undergraduate, expected grade, prior achievement, or personality); teaching conditions (e.g., class size, elective/required, discipline/department, workload/difficulty, course level, or time of day); instructor characteristics (e.g., gender, academic rank, age, research productivity, or teaching

level); content of the instrument (e.g., unidimensional versus multidimensional); and procedural factors (e.g., purpose of ratings, anonymity of raters, presence of instructor, timing of administration, format, sampling, or leniency/severity of the rater). In discussing variables that do and do not have the potential to bias student ratings, Cashin (1995) categorizes them according to three subdivisions: instructor variables (age, gender, teaching experience, etc.), student variables (personality, age, gender, etc.), and other variables (course size, course time of day, distribution of evaluation surveys, etc.). With the exception of the catch-all term “other,” this approach to classifying the literature related to discriminant literature provides a simple approach for categorizing the literature. Rather than using the term “other,” however, this study uses the term “course variables.”

Since this study classifies the issue of student integrity in completing SRTs as a potentially biasing factor, that consideration calls for a more comprehensive review of literature specifically dealing with discriminant validity. Course variables researchers have considered as having the potential bias include course level, class size, workload/difficulty, factors related to administration and stated purpose of the ratings, and academic discipline. Student variables having the potential for bias include prior subject interest, reason for taking the course, student level, instructor gender, motivation, consumerism, feelings of indebtedness, retribution, fear of retaliation, instructor reputation, and halo effect. Finally, the factors of rank, years of experience, type of instructions, grading leniency/expected grade, instructor personality, enthusiasm (the “Dr. Fox” effect), and instructor likeability (or rapport) comprise the instructor variables. Keep in mind that the literature related to these various factors focuses on one or two of these factors and does not label them as belonging to instructor variables or student variables. Indeed, in some cases one might argue that

instructor likeability, for instance, involves student perception (a student factor) as much as it would involve some conscious action or behavior on the instructor's part. So the placement of certain items under certain categories is merely a means of categorizing the literature for the purposes of this study. Some studies indicate that certain variables show little or no relationship to ratings in regards to teaching effectiveness whereas other studies are still questioning the potential bias of a few.

Discriminant Validity (Bias) Studies Related to Course Variables

Course variables concern factors that pertain to the nature of the course itself as opposed to factors relating more directly to students or faculty members. These course variables include course level, class size, workload/difficulty, factors related to administration and stated purpose of the ratings, and academic discipline.

Non-Bias Factors.

Despite the report from Cashin (1995) that course size, the way in which surveys are distributed, or the time of day for the course have little or no effect as biasing factors in rating teaching effectiveness, only the time of day appears consistently in the literature as a non-factor (Aleamoni, 1987; Feldman, 1978).

Course Level.

In a study by Wigington, Tollefson, and Rodriguez (1989), in which they examine possible correlations between three variables related to the class (type, level, and size) and three variables related to the instructor (reputation, rank, and sex), they find that, overall, instructors of lower-division courses received lower ratings than instructors of upper-division or graduate courses, with no difference in ratings for females or males. However, as will be discussed later, male instructors of upper-division courses received higher ratings than female

instructors. The significance is great enough that the authors recommend that institutions include the level of the course when interpreting the results of student ratings.

Various studies have found indications that course ratings may vary according to level. Marsh (1980, 1983), while conducting SEEQ research, finds that students give higher ratings to higher level courses, particularly graduate courses. Additionally, according to a meta-analysis by Feldman (1978), student ratings relate positively to course level although the effect diminishes when other background variables are controlled.

Class Size.

With the factor of class size, the question is whether students' ratings accurately reflect that class size affects instructors and their instruction, i.e., non-bias, or whether class size merely affects students' attitudes toward the course and their perceptions of their instructors to the point that ratings fail to reflect the instruction that the students receive, i.e., a bias (Feldman, 1984). Faculty members often contend that the size of the course affects their ratings, with the larger the class, the lower the ratings. The meta-analysis that Feldman (1984) conducted, however, reports mixed results. Although only 2 studies found a positive correlation between class size and student ratings compared to 22 studies finding an inverse relationship between class size and ratings, another 22 studies found no relationship with 12 more reporting a curvilinear relationship in which both small and large classes received high ratings while medium-sized classes received the low ratings.

Most research related to course size indicates that it creates a slight effect at most. Both Costin, Greenough, and Menges (1971) and Kulik and McKeachie (1975) discuss various older studies that provide the same mixed results that Feldman also reported, with some suggesting large classes receive lower ratings while others find that students gave high ratings

in overall teaching ability to instructors of large classes. Additionally, they mention studies that student ratings may vary according to a particular aspect of teaching performance in relation to class size. For example, instructor preparation received a high rating in large classes but ratings for the amount of original thinking students thought is demanded of them was lower as well as the finding that smaller classes generate a stronger feeling of goodwill between student and instructor but that larger classes make more efficient use of class time. Although Aleamoni (1987) lists class size as one of the myths that faculty members often hold about student ratings, in fact, the results still appear to be mixed.

Class size correlated moderately with the dimensions of Group Interaction and Individual Rapport on the SEEQ but not with other SEEQ dimensions or with the overall course or instructor ratings (Marsh, 1980, 1983). Because class size relates only to these two dimensions, Marsh uses it as an example of the importance of examining the multi-dimensionality of students' ratings, and he also considers it not to be a biasing factor for ratings of teaching effectiveness. In an earlier study, Centra and Creech (1976) discover a curvilinear effect, with the smaller and larger classes receiving the highest ratings and those classes containing between 35-100 students receiving the lowest. Feldman (1978, 1984) and Costin, Greenough, and Menges (1971) also cite additional studies that report this curvilinear effect.

More recently, Mateo and Fernandez (1996) find that although class size does have some impact on teaching ratings, this relationship differs somewhat as a function of teaching quality and motivational and interactive skills, and the effect sizes are quite small. They conclude that classes with fewer than 10 students seem best for obtaining teaching excellence, but with more than 10 students in lecture-type courses, no effects worthy of

mention appear for other class sizes.

Conversely, a study by Baker and Copp (1997) indicates that class size may have an impact on women professors' ability to fulfill students' gendered expectations that women professors are more sympathetic, concerned, and lenient than their male colleagues. Their study demonstrated that when Baker, a pregnant professor teaching a course entitled "Women, Men, and Society" from a feminist perspective, taught a 23-person summer section of this class, the classroom conditions (teaching a class containing fewer students) helped minimize an authoritarian professor role because Baker was better able to have individual contact with the students, thus reducing the likelihood of contradictions between students' expectations for women and for professors, and the conditions also led students to perceive Baker's pregnancy in a different and positive light. Although their main contention is that students' expectations of women and of professors impact student ratings, classroom size can affect how easily women professors can meet student expectations, thus impacting SRTs.

A study by Wigington, Tollefson, and Rodriguez (1989) finds that female instructors received higher ratings when teaching small (25 or fewer students) or intermediate-size (26-49 students) classes than male instructors whereas males received higher ratings when teaching large (50+ students) courses although no difference between the ratings of female and male instructors occurred for those teaching lower-division courses. Yet male instructors teaching upper-division courses received higher ratings than female instructors. These findings support Baker and Copp's conclusions related to a gender bias based on student expectation. (See the discussion of "Instructor Gender" in the studies related to student variables later in this chapter.)

At best, the results related to course size are mixed. Feldman (1978, 1984) also offers

the idea that at least two possible interpretations exist to explain the results. One is that either the course characteristic directly biases students' reactions to the instructor. In other words, class size affects how students rate instructors. The other is that it creates certain conditions that influence the instructors' behaviors and effectiveness. In other words, instructors may teach more effectively and put more effort into teaching smaller classes than larger ones, in which case, the ratings would not represent a bias but rather a reaction to instruction.

Workload/Difficulty.

A common belief among some faculty is that students rate "easier" courses more favorably and penalize the "more difficult" courses, thus making the workload or difficulty of the course a potential bias. In fact, the workload/difficulty effect on students' ratings was one of the largest Marsh (1980, 1983) finds in his studies of the SEEQ. But Marsh reports that workload/difficulty is positively correlated with ratings, the opposite direction to that predicted as a potential bias; therefore, workload/difficulty appears not to be a bias to student ratings. Marsh (1982a) reconfirms these findings in another study that relates differences in the ratings of the same instructor teaching the same course on two different occasions to differences in the background characteristics of the two courses. A study by Freedman and Stumpf (1978) also presents results that more difficult courses being associated with more favorable ratings while Frey, Leonard, and Beatty (1975) find no significant correlation between workload and student results of their final exams.

In their study using the student perceptions of teaching effectiveness (SPTE), Burdsal and Bardo (1986) determine that student ratings of teaching quality are unrelated to student perceptions of course difficulty. According to their findings, when they correlate expected

grade with course difficulty but not with overall quality and when they produce a summated scale of “teaching quality,” which includes items from both the overall quality factor and the course difficulty factor, a correlation appears between the expected grade and “teaching quality.” However, they assert that this finding will result in “inappropriate and erroneous conclusions” because one can spuriously interpret extraneous variables related only to course difficulty to predict overall teaching quality. Feldman’s meta-analysis of dimensions of instruction and student achievement (1990) also concludes that no association exists between student achievement and the difficulty or workload of the course.

Administration and Stated Purpose of the Ratings.

How and why instructors and/or institutions administer student ratings have been the focus of some studies of potential biasing factors. Aspects studied include anonymity of student raters, the purpose of the ratings, and the time in the term when the instructor administers the ratings.

In terms of anonymity, Feldman (1979) reports that anonymous ratings tend to be somewhat lower than non-anonymous ratings but that this result can vary with other circumstances. Braskamp, Brandenburg, and Ory (1984) note similar findings but recommend anonymous evaluations. According to a study that Dwinell and Higbee (1993) conducted examining students attitudes towards ratings, students feel confident that the method used for conducting student ratings protects their anonymity. The approach involves instructors reading standard instructions to the students and then leaving the room while the student complete the ratings with one student then taking the results to a given collection point. Just the fact that the study investigated student perceptions about anonymity implies that if students feel their anonymity is not protected, that factor could affect how they

respond.

When ratings are intended for administrative purposes rather than for faculty feedback or for research purposes, ratings tend to be higher, according to Feldman (1979). In a review of studies concerning halo effect (a distortion in judgment resulting from overall impressions having an undue effect on specific judgments), Murphy, Jako, and Anhalt (1993) report that research on the effects of the purpose of rating suggests that ratings tend to be inflated and tend to minimize within-ratee distinctions and maximize distinctions between ratees, thus supporting Feldman's finding. (A more extensive discussion of halo effect appears later in this chapter.) Frankhouser (1984), however, finds that stated purpose has no effect on global ratings. Such a result would support those who favor the use of global ratings as adequate to evaluate instructors and courses although, overall, purpose of the ratings appears not to constitute a biasing factor. Dwinell and Higbee (1993) contend that what an instructor tells students about the purpose of the ratings may affect how students respond to the ratings items, and therefore they advocate having instructors read a standardized explanation that includes an explanation of how instructors and the institution will use the ratings. The implication is that an instructor could bias how students rate them based simply on this factor alone. For example, adjunct faculty members who tell students that the student ratings could affect their continued employment, students might rate them higher (or even lower) than they deserve, depending on how much credibility they give what the instructors are telling them. Since this would be a factor that is external to actual instruction and its effect on student achievement, it would constitute a bias.

Feldman (1979) reports that the time when ratings are collected (middle of term, end of term, during the final exam, after course completion) tend to be similar. Braskamp et al.

(1984) suggested that ratings collected during the final exam may be lower and that ratings collected at midterm may be unreliable if the instructor can identify students.

Academic Discipline.

Researchers have considered the possibility that students rate courses and instructors within certain discipline areas differently because of the nature of the discipline. For example, students, on average, may rate courses in the natural sciences, which are generally highly structured, differently than they rate humanities courses, which may be more loosely structured and rely more on discussion.

According to Feldman (1978), who reviews studies comparing ratings across disciplines, ratings for English, humanities, arts, languages, and perhaps education are somewhat higher than average compared with social sciences, physical sciences, mathematics, engineering, and business administration, which tend to receive somewhat lower than average ratings. Furthermore, Gigliotti and Buchtel (1990) find that courses containing math content produce lower judgments of fairness, less satisfaction, and more negative effects on self-feelings than courses containing no math although class type did not impact the pure evaluation dimensions.

Discriminant Validity (Bias) Studies Related to Student Variables

Student variables involve factors that students seem more directly related to students and their reactions. Those potential biases related to student variables include prior subject interest, reason for taking the course, student level, instructor gender, motivation, consumerism, feelings of indebtedness, retribution, fear of retaliation, instructor reputation, and halo effect.

Non-Bias Factors.

Variables related to student factors that have little or no effect as biases for rating teaching effectiveness include personality, age, gender, level (Cashin, 1995) or grade point average (Cashin, 1995; Feldman, 1976a; Freedman & Stumpf, 1978). Freedman and Stumpf (1978) also contend that the learning style of students, using the Learning Style Inventory, to have no biasing effect on student ratings.

Prior Subject Interest.

Marsh (1980, 1983) examines the effect of prior subject interest by students on SEEQ ratings and determines it to be highly correlated with the Learning/ Value dimension. Increased student interest in the subject apparently generates a more favorable learning environment and enhances effective teaching, then carrying over into favorable student ratings. Because prior subject interest influences ratings in a way validly reflected in student ratings, the influence should not be interpreted as a bias to student ratings according to Marsh (1982a, 1987). However, since prior subject interest is more a function of the course rather than the instructor, Marsh cautions that prior subject interest may represent a source of “unfairness” when institutions use ratings for personnel decisions.

Although not stated in terms of “prior” subject interest, Doyle and Whitely (1974) report that students’ liking for the subject (which could indicate a prior student interest) correlates with ratings of how much they learned, the motivational effectiveness of the instructor, and overall course and instructor effectiveness. These findings appear to support Marsh’s contention that prior interest is naturally connected with favorable ratings related to learning. Common sense also supports this conclusion, and the fact that Doyle and Whitely, as mentioned earlier, find no correlation between students’ liking the subject matter and their

liking for the instructor as a person points to prior interest not representing a biasing factor as Marsh contends.

Reason for Taking the Course.

The belief underlying this bias is that students who choose to take a particular course, as opposed taking the course as a requirement, are more likely to rate it and the instructor more highly than those required to take the course. Therefore, rating forms often include the reason a student is taking the course as one of the items. According to Marsh (1987), large empirical studies by Brandenburg, Slindle, and Batista (1977), Pohlman (1975), and Centra and Creech (1976) have indicated a positive correlation of a course's electivity to student ratings. Costin, Greenough, and Menges (1971) cite earlier studies showing required courses receive slightly lower ratings than elective courses. To account for seniors rating instructors more highly than "less experienced" students, Frey, Leonard, and Beatty (1975) present the idea that seniors might be taking a course as an elective rather than as a requirement as one hypothesis for the higher ratings. On the other hand, Aleamoni (1987) reports mixed results and contends that higher ratings resulting from students taking a course as an elective rather than a requirement is a myth, and Freedman and Stumpf (1978) find no significance between the reason for taking the course and student ratings.

Marsh (1980) claims that the dichotomy of a course being elective or required is too simplistic. Therefore, the SEEQ allows students to indicate one of six course designations: (1) major requirement; (2) major elective; (3) general interest; (4) general education requirement; (5) minor/related field; and (6) other. In studying the impact of these reasons for taking a course on student ratings using the SEEQ, Marsh discovers that all SEEQ factors tend to correlate positively with the percentage of students taking a course for general interest

and as a major elective and negatively with the percentage of students taking a course as a major or a general education requirement. But after controlling for the 16 other background characteristics, general interest emerges as the only reason to have a substantial effect on ratings. Therefore, the results of those students taking a course as an elective (for general interest) are more likely to show a bias.

Motivation.

Like other potential biases, motivation is one that may or may not actually be a bias depending on how one views it. It stands to reason that if students enter a course with a high degree of motivation to begin with, either because of personal interest in the subject, a personal choice of taking a course from a particular instructor, or a combination of these and perhaps other factors, one may assume that this motivation could color their overall rating of the course and/or the instructor. However, motivation might also result from something the instructor does as part of the course to inspire student learning. As the results of the meta-analysis that Feldman (1990) conducted on the dimensions of instructions and their impact on student achievement revealed, “high standards and motivation” is one of six dimensions that account for 10% of the variance in student achievement. Student motivation that the students bring into the course would be an external (or discriminant) factor, i.e., a bias, whereas the motivation that an instructor would inspire through her or his teaching would be an internal (or convergent) factor directly related to teaching and would therefore not be a bias. However, Feldman also cites Leventhal (1975), who points out that a third variable such as motivation might independently affect both teacher performance and student learning, thus accounting for the correlations between instructional characteristics and student achievement.

Student Level.

In their review Costin, Greenough, and Menges (1971) cite earlier studies showing general agreement between undergraduates' and graduate students' ratings of the same instructors as well as studies finding that upper division students tend to rate instructors more favorably than do students in lower divisions as well as a study that graduate students assign higher ratings than undergraduates. However, one might argue that upper division and graduate students are more likely enrolled in courses in their major field of study or are courses that students elected to take, thus accounting for the higher ratings. Costin et al. also cite studies showing that undergraduate and graduate students taking the same courses with the same instructors show no significant differences in their ratings.

Kulik and McKeachie (1975) present a slightly different twist on the idea of student level in relation to ability level rather than class level. They report the findings of a 1950 study by Elliott in which he discovers that students respond according to their ability levels in relation to whose ability level the instructor teaches to. "Bright students" rate those instructors who teach to their level more highly than instructors who teach to the level of "low-ability" students. Conversely, the instructors who teach to the level of the low-ability students receive higher ratings from those students.

A more recent study by Spencer and Schmelkin (2002) finds significant and meaningful differences due to level in relation to importance of indicators of effective teaching. All means are relatively high, yet the higher means for sophomores (4.42) and juniors (4.52) and slightly lower ones for seniors (4.28) and graduate students (4.26) lead the authors to conclude that it appears that the more exposure students have to teaching, the less value they give to the importance of indicators of effective teaching. Although they do not speculate

about these findings in relation to their implications for student ratings, these results can imply a reason for any increase in ratings coming from seniors and graduate students, who may be less critical of instructors, perhaps in correlation to becoming more independent learners as well as perhaps reflecting a higher degree of self-assurance. However, this factor appears to be one that needs more research to confirm these speculations.

Instructor Gender.

If the mere fact of the teacher's gender determines in whole or in part students' ratings of teaching, this constitutes a bias. However, if actual dispositions, practices, and effectiveness determine the rates, then the ratings will be valid rather than biased (Feldman, 1993). Studies related to gender constituting a biasing factor are mixed. Because of the mixed results, Aleamoni (1987) considers the belief that instructor gender affects student ratings to be a myth. In their early review of older literature, Costin, Greenough, and Menges (1971) cite studies providing mixed results, some finding no significant differences that male and female students make in rating instructors nor in the ratings male or female instructors receive; one by Bendig in 1952 reporting female students being slightly more critical of their male instructors than male students; a 1969 study by Walker showing that female students rated female instructors significantly higher than they rated male instructors; and McKeachie, Lin, and Mann's 1971 study reporting that female students tend to rate instructors higher who also rate high on "skill" and "structure." In more recent works, Feldman (1993, 1997) contends that gender of the instructor does not bias students' ratings based on his meta-analysis of the literature. In considering overall evaluations of male and female instructors, across 28 studies the average correlation between instructor gender and students' overall ratings is .02, indicating that female faculty members received a slightly higher overall

evaluation than did males. As Feldman points out, on average gender explains only 4/100 of 1 percent of the variance in overall ratings. However, some studies provide some evidence indicating that students may rate same-gendered instructors a little more highly than opposite-gendered instructors (Feldman, 1997).

Although it may be the case that some students have a gender bias toward faculty of the opposite sex, studies examining instructor gender consider gendered expectations that some students may have of faculty. According to Baker and Copp (1997), students may hold contradictory and unrealistic expectations of female professors because of women's culturally defined gender status clashing with their occupational status, making it hard for women faculty members to receive outstanding teaching evaluations. They contend that students judge women by their gender performance first and their teaching second. In their study, they collect and analyze the SRTs of the students who took a course entitled "Women, Men, and Society," which Baker taught from a feminist perspective. Baker offered the course in the spring, summer, and fall semesters. She taught the course using the same materials, tests, and textbook. The only differences were that the spring and fall classes were large lecture courses (200 students each) while the summer class had only 23 students, and Baker became pregnant during the spring semester so that she was noticeably pregnant during the fall semester.

They claim that students have gendered expectations, expecting women professors to be more sympathetic, concerned, and lenient than men. "Gender plays a crucial and sometimes deleterious role in the student-female professor relationship" (p. 30). Results from their study show that when the professor fulfilled students' gendered expectations during the summer class, they rated her teaching highly despite her being pregnant and teaching from a feminist

perspective. But when she did not meet their gendered expectations during the fall semester, their ratings reflected a lack of tolerance for her position and her pregnancy. Combining gendered expectations, feminist course content, and pregnancy could cause strong reactions as students “give their professor negative evaluations as a way to punish her for violating their gendered expectations” (p. 30). Also, they assert that the more women and men professors adhere to gender-appropriate teaching styles, the higher the ratings. One might argue that a caring attitude toward students may produce higher student ratings regardless of gender and that a smaller class size promotes more frequent and better student-instructor interaction, which may also account for higher student ratings. Their study is also limited to one instructor and three classes. However, the study does raise the possibility that, indeed, some students bring gendered expectations to their classes and that these expectations bias student ratings. Feldman’s meta-analysis (1993) also concedes that differences in students’ expectations and demands male and female instructors may “underlie similar ratings that students have given on evaluation forms just as then may underlie differences in ratings” (p. 178). But he also contends that what may appear true in one college may fail to appear in others.

Similar findings of gender expectations that may bias results appears in a study Bennett (1982) conducted; however, Bennett’s findings are mixed. Although she found that students have similar standards of reference for male and female instructors, students rate female instructors as being warmer, more supportive, and more personally charismatic than male instructors. Additionally, perceptual ratings prove to be equally consequential for men and women, particularly in terms of being perceived as warm, encouraging, and nonauthoritarian. However, students do receive greater time and personal support from female instructors yet

seemingly hold women to a stricter standard of accessibility. In terms of student expectations of accessibility, Bennett concludes that students evaluate female instructors negatively when they fail to meet a gender expectation for warmth and accessibility as well as reward them with higher ratings when they do.

Martin (1984) also obtains mixed results from her study of 240 female and 154 male students at a large Midwestern university. She finds that both female and male students evaluate male and female social science faculty and female women's studies faculty the same. Only with males evaluating female social science faculty do signs of sex bias appear. The results of the study reveal that male students rate females who displayed both "feminine" attributes (friendliness, smiling often, and making frequent eye contact) and "masculine" attributes (confidence and decisiveness) most highly. These results lead Martin to conclude that male students likely judge the competence of female instructors on the basis of personal characteristics typically associated with female behavior. Thus a degree of sex bias is at work in the evaluations of female faculty.

Similarly mixed results appear in a study that Rubin (1980) conducted. In the first of her investigations, she sought to determine whether students would perceive the ideal professor differently depending on the gender of the professor. Results indicate that students appear more concerned with the role of the professor than the sex of the person in that role. With her second aim, however, which is to determine whether students use different terms when addressing their male and female professors, she finds that students tend to address female professors in the 26-33 age group using their first names while they address males in the same age group more frequently using some sort of title and their last names. In particular, female students consistently use the familiar address with their female professors

while using the title “professor” along with the last name for their male professors. It is unclear, however, whether this result occurs because younger female professors lack the status of their male counterparts or because female professors may be easier for students, especially female students, to identify with.

As mentioned earlier, the study that Wigington, Tollefson, and Rodriguez (1989) conducted determines that although no difference occurs between the ratings of female or male instructors teaching lower-division courses, students rate males higher than females for the upper-division courses. Additionally, males receive higher ratings in large classes (50+ students) whereas females receive higher ratings in small and intermediate-sized classes. In examining the correlation between rank by sex, they discover that no difference between male and female instructor ratings occurs for assistant and associate professors, but for graduate teaching assistants and professors, ratings are higher for females than for males. The results showing a class-size difference by gender could support Baker and Copp’s (1997) findings and may reinforce their contention that students in smaller classes may perceive female instructors as meeting their gendered expectations of being more sympathetic, concerned, and lenient. Wigington et al. (1989) explain the difference as being the result of smaller classes being more conducive to expressiveness.

To the extent that preconceptions of female and male instructor behavior affects student expectations and to the extent that these preconceptions influence student ratings of instructors according to factors other than effective teaching, then these would qualify as biases based on gender. What seems confounded is that in some respects females rate more highly than males whereas in others the reverse is true.

Consumerism.

The concept of “consumer satisfaction” is not new. Costin, Greenough, and Menges (1971) cite studies from McKeachie in 1969, in which McKeachie uses the term “consumers” of a course in reference to students. The notion of consumerism appears to be inextricably bound with the view that somehow instructors must become “entertainers” or easy graders to please these student consumers. In their review, Costin et al. (1971) report that early studies conducted as early as 1934 and dating to 1960, find little direct evidence to support the view that students rate entertainment value more highly than information value. They mention a study by Guthrie conducted in 1954, determining that instructors receiving the highest student ratings are “substance teachers” and not merely entertainers.

Recent essays (Bellah, 1999; Delucchi & Smith, 1997; Long & Lake, 1996) take aim at the concept of consumerism in relation to the deleterious effects this concept has on higher education. These effects include undermining professorial authority, causing professors to be reluctant to hold students to exacting standards of performance and encouraging passivity among students, excusing them from taking an active part in learning, and leading them to feel entitled to passing grades (Long & Lake, 1996), and the effects may also extend to grading leniency and to faculty wanting to be liked (Delucchi & Smith, 1997).

Cruse (1987) distinguishes between evaluation models aimed at teaching the student, which he labels as the “Learner Model,” and those aimed at pleasing the student, dubbed the “Consumer Model.” He contends that a learner model is not likely to require or need student ratings of teachers, presumably because student performance will provide the only measure of teaching effectiveness needed. For Cruse, the Consumer Model tends to promote “the

salesperson, the popularity poll” but with the result that by doing so the institution may “maximize the popularity features of the ratings to promote enthusiasm and sensitivity to student attitudes” (p. 734). Cruse asserts, in effect, a series of causal relationships, with student ratings promoting a consumer model, which, in turn, encourages instructors to develop “charismatic talents” (i.e., stresses likeability), which then result in higher ratings.

Delucchi (2000) concludes that, “The negative effect of likeability on perceived learning, juxtaposed with its importance in positively predicting global SETs gives credence to concerns about student consumerism in higher education” (p. 228). Bellah (1999) attributes American concern with “the bottom line” as a contributing factor driving academic decisions as well as causing students with a consumer mentality to have difficulty accepting institutional or professorial authority as well as expecting high grades simply because they have paid for them. Perhaps this consumer mentality is partially responsible for the “culture of disengagement” that Flacks and Thomas (1998) refer to. They mention how faculty members feel pressure to “dumb down courses” because students fail to read the assignments, attend class regularly, or engage in rude behavior in the classroom. Martinson (2000) also expresses concern about students, parents, administrators, politicians, and others who define good teaching in terms of complying with student preferences as opposed to doing what is in the students’ best interest. Therefore, this factor of consumerism may impact other factors that, in turn, impact SRTs. As Delucchi (2000) contends, the impact may be in terms of the pressure some instructors may feel that they need to cause their students to like them (and to show that liking through higher student ratings), thus leading them to lower their workload. Gillmore and Greenwald (1999) theorize that in response to student evaluations that complain of too much reading or insufficient time being devoted to difficult material,

instructors will modify their syllabi (i.e., reduce the workload), thus leading to indirectly participating in grade leniency, which results in students obtaining higher grades.

Consumerism as a factor, therefore, appears intertwined with other possible biases such as likeability/rapport, workload, and grading leniency/grade expectations. However, as is true with so many of the other potential biasing factors, the studies provide no clear evidence to determine whether higher student ratings result simply from instructors becoming entertainers or lowering their grading standards to pander to students or whether what some may perceive as indulging student consumerism may, in fact, reflect a trait that leads to more effective student learning and consequently higher student ratings.

Feelings of Indebtedness.

El-Hassan (1995) finds that students give higher ratings to more experienced instructors although it is unclear whether these ratings are the consequence of higher learning with more experienced teachers or because the students believe that more experienced teachers are better or that they feel indebted to them and therefore they rate them more highly. If the ratings are due to a feeling of indebtedness, this would constitute a biasing factor.

Retribution.

Just as some students may feel indebted to some instructors, therefore rewarding them with higher evaluations than deserved, the opposite reaction also likely occurs for some students. Studies have indicated that some students may use student ratings as a means of gaining retribution, particularly if the students feel as if they are receiving lower grades than they deserve. Costin, Greenough, and Menges (1971) refer to this phenomenon as “student hostility”; however, their review of earlier studies finds no studies providing any direct

evidence to support student retribution. Whether students' taking retribution extends to faculty members who hold students to higher grading standards is unclear. As discussed in connection with the grading leniency factor, certainly some faculty members believe that easy graders receive higher ratings, the implication being that hard graders receive lower ratings, thus resulting in the belief that students use ratings as a form of retribution.

Holmes (1972), as discussed earlier, finds retribution to be the case in his study related to grading bias. Furthermore, Martin (1984) in her study of possible sex bias in student ratings connects the idea of retribution for faculty who violate acceptable role behavior based on gender. "Since it is reasonably certain that revenge plays a part in at least some low student evaluations, we may conjecture that it plays a larger part in the evaluations of those faculty who upset students by violating zones of acceptance" (p. 487). Martin adds that because of conflicts between established sex roles and traditional standards for expressing authority, female faculty may violate these norms more easily than their male counterparts.

The factors of both indebtedness and retribution are ones that appear more directly related to a conscious decision on the part of students to reward or punish certain instructors through faculty ratings. To the extent that they actually occur, they both relate directly to the issue of student sincerity in completing ratings. However, another aspect of retribution concerns the fear of retaliation on the part of the faculty member.

Fear of Retaliation.

Apparently, no specific studies related to student fears of instructor retaliation have been conducted although the literature sometimes mentions the importance of having a neutral person administer the ratings to remove, or at least mitigate, the fear that some

students have. Miller (1988), for example, comments about students' being suspicious of completing evaluation forms that an instructor hands out in class and fearing that the instructor will recognize the source of any critical comments and may be vindictive towards any critics of his or her teaching. Additionally, various sources call for the need for student ratings being anonymous to encourage open student responses (Costin, Greenough, & Menges, 1971) or encourage results to be withheld from instructors until after the instructors submit their final grades (Braskamp & Ory, 1994). The implication of such recommendations is that some sort of fear that any negative ratings on the part of students could result in the instructor lowering their grades should the instructor know which students provided low ratings, thus resulting in students giving less candid or sincere ratings and/or comments.

As part of their study on student perspectives on teaching and its evaluation, Spencer and Schmelkin (2002) present findings that, overall, students show little fear of retaliation. However, they also indicate that of those students registering concerns with retaliation are also those who are more reticent about completing ratings. These results reveal that although fear of retaliation likely does not bias SRTs in general, this factor may affect the integrity of the responses of students who do have this concern.

Apathy.

Since the preponderance of studies related to SRTs investigate the rating instruments, it may be understandable why student lack of concern, or apathy, fails to appear in the literature. As studies begin to shift more attention to students, possibly more research will focus on the impact that apathy exerts on students and how much effort and thought they put

into ratings. In one recent study, Spencer and Schmelkin (2002) report that despite not being reluctant to complete SRTs and not having any particular fear of bias, students are unsure whether their opinions matter as well as to what purpose the ratings are put. As a result, “they may not pay much attention to them” (p. 406). This finding leads the authors to contend that the problem becomes one of convincing students that their opinions do matter. The impact that apathy may have in terms of biasing results clearly is one that requires further research.

Instructor Reputation.

Wigington, Tollefson, and Rodriguez (1989) focus on the individual student evaluation and examine the effects of three class variables (size, type, and level) and three instructor variables (rank, reputation, and sex). They find that the most powerful independent variable is instructor reputation. According to their results, faculty members for whom students report taking the course because of the instructor’s reputation achieve higher mean ratings than other instructors. Additionally, they find that full professors receive the highest mean rating when they combined reputation with rank as well as that full professors also receive the lowest mean ratings when reputation is not a factor in students’ decisions to take a course. However, reputation is not a significant factor in connection with the level of the course (lower division, upper division, or graduate), the gender of the instructor, or course size.

As Wigington et al. indicate in their discussion, critics of student ratings contend that these findings support the Dr. Fox studies in which instructors receive higher ratings based only on delivery rather than content as well as add fuel to belief that student ratings are merely popularity contests. However, the authors conclude that those faculty members who

have developed positive reputations as effective teachers earn higher ratings and that these consistently higher ratings based on faculty reputation reflect well-developed teaching skills.

Halo Effect.

“Halo effect is the tendency of a rater to form an overall positive impression of the person being rated, and on this basis assign indiscriminately high ratings to all dimensions of performance, including those that, in reality, are statistically independent” (Murray, 1997, p. 178). Marsh (1987) characterizes a halo effect as a generalization from some subjective feeling that is unrelated with effective teaching, an external influence, or an idiosyncratic response mode. When the same raters evaluate both specific teacher characteristics and overall teaching effectiveness, inflation of process-product correlations due to halo effect is presumably more likely to occur. Solomonson and Lance (1997) present Nisbett and Wilson’s (1977) definition, which characterizes the halo effect as “the influence of the rater’s general impression on ratings of specific ratee qualities” (p. 665). Thus, most researchers view halo effect as a form of rater bias. According to Cruse (1987), “the halo means that teachers may not be able to rely on the ratings as useful and valid opinions of separate skills and as guides to self improvement” (p. 726). However, as Becker and Cardy (1986) point out, citing Harvey, “...the notion of halo assumes that general impressions yield unduly high intercorrelations among performance dimensions or low variance across dimensions” (p. 669) but that no one has provided empirical support for these assumptions.

As mentioned earlier, the results of the study that Williams and Ceci (1997) conducted indicate that, indeed, a change in a single factor (instructor enthusiasm) results in higher ratings for all factors, not just the enthusiastic factor in the rating instrument. In his meta-

analysis of studies of students' views of superior instructors, Feldman (1976b) also acknowledges that once students have an initial impression that an instructor is superior, for whatever reason, a halo effect could occur, resulting in higher ratings on specific aspects of teaching. But he draws conclusions opposite from Williams and Ceci's. Interestingly, Feldman lists enthusiasm for teaching the subject as one of the factors that could receive a higher rating because of a halo effect rather than as causing a halo effect. These increases in all factors demonstrate a possible halo effect. Also, Powell (1977) notes that his findings suggest that students form an overall impression of the instructor and then respond to questions on the rating form partially on the basis of their global impression. Moreover, he asserts that the extent that the impact of the global impression affects the response appears to increase with the subjectivity of the question although the global impression influences even objective questions.

In his study comparing student ratings with instructor self-evaluations using the same instrument, Marsh (1982b) discovers that at least some halo effect occurs in the student ratings but not the instructor ratings. However, Marsh also offers two mitigating factors to explain the halo: (1) the student ratings are more reliable (median coefficient alpha = .94) than the instructor self-evaluations (median coefficient alpha = .82), thus inflating the correlations among student evaluations relative to the convergent validity coefficients; and (2) the possibility of true relationships among the rating factors exists, i.e., the instructional evaluation factors may be correlated resulting in a relationship that is independent of method. Also noting the possibility of a halo effect are Cashin, Downey, and Sixbury (1994) as they suspect that "students are reluctant to give any low ratings to an instructor whom they perceive as effective, even though a particular item does not apply to his or her course" (p.

650). Braskamp, Ory, and Pieper (1981) conducted a study comparing student ratings given on objective rating forms with responses to open-ended questions and learn that students are more consistent in their ratings of instructors using the two methods than in their ratings of the course. They acknowledge that the consistency between the two ratings may result from a halo effect, but it may also result because instructor characteristics are more observable than are course characteristics, thus resulting in the higher consistency. Besides research that finds that halo effect happens or may happen, other research has explored the causes of halo.

Fisicaro and Vance (1994) categorize conceptual definitions of halo effect according to the implied or stated cause: “inadequate discrimination of ratee behavior, a general impression of the ratee, or perception of the ratee in terms of a salient dimension or characteristic” (p. 366). In another study, Lance, LaPointe, and Stewart (1994) test these three causal models of halo rater error. Their study investigates whether they can manipulate rating contexts that will favor the operation of inadequate discrimination, salient dimension, or general impression models of halo error. They learn that despite their manipulations designed to produce the three different forms, a general impression (GI) model of halo rater error best accounts for halo rating error. “This suggests that, rather than being context dependent, halo error may occur in the form of GI halo error regardless of rating context” (p. 338). Lance et al., caution that more research is needed to support their results. In effect, however, these results support Powell (1977) as well as Cashin et al (1994). Once students have formed a general impression of instructors, these impressions may then color all the ratings, especially global ratings. Also, if one factor such as instructor enthusiasm generates this general impression, as Williams and Ceci (1997) contend, then these results may argue for using multidimensional ratings that have a better chance of discriminating between

various factors.

Information-processing models, according to d'Apollonia and Abrami (1997), posit that individuals use common prototypes to organize their schemas of persons, roles, and events. Items on ratings forms can trigger the raters' associative network, in which they have stored general impressions. Fiske and Neuberg (1990) contend that impression formation functions on a continuum ranging from category-based impressions, i.e., stereotyping, to fully individuating impressions using attribute-oriented processes. One of the premises they investigate asserts that people will first attempt to fit their impressions into an existing category, and it is when characteristics do not easily fit into a category that people will move toward more attribute-oriented processes. The implication of their model for student ratings is that for those students who have already categorized or stereotyped the instructor, it is more likely that halo effect will occur. However, their model also presents factors that have a mitigating effect on people simply categorizing. They show that the more attention people pay, the more likely people will move toward attribute information. Furthermore, motivation influences impression formation. This motivation may come from a third party, may be the target, or may be the perceiver's self. Although Fiske and Neuberg do not relate their model specifically to student ratings, one can infer situations that could relate to these motivators. For example, an administrator or other third person who comes into a classroom and explains the purpose and importance of faculty ratings could provide students with a motive to take their time, to pay closer attention to each question, and to provide ratings based more on specific instructor attributes rather than their general impressions of the instructor that, in turn, might lead to the appearance of halo effect. Similarly, if students know that the "target" of student ratings is to help instructors improve teaching and that the instructor(s) and the

administration will take the ratings seriously, then this could trigger paying closer attention to attributes. Finally, individuals whose personal values require them to try to be objective and not overgeneralize would have the motivation to use attributes rather than categories during ratings. Complicating impression formation even more, their model asserts that attentional and interpretive responses to attributes mediate motivational influences. To explore the area of impression formation theory beyond this brief overview is beyond the scope of this study. However, what this work indicates is that it is not only possible but also likely that some students form a categorical impression of an instructor and that this stereotype then manifests itself in halo effect on student ratings. Yet it also indicates that other students will pay more attention to individual attributes and will be more discriminating in their responses to specific questions in the rating instrument. Moreover, the implication is that factors can prompt students to be more attribute oriented, thus reducing the chance of halo error biasing the results of the ratings. In general, the model implies that halo effect may or may not occur for any given student or within any given class, depending on a variety of factors; in other words, bias may or may not occur.

Expected grade is a factor that some have considered as creating a halo effect in terms of the overall ratings. However, McKeachie (1997) considers the grade-related halo effect in judging course characteristics argument to be irrelevant to the validity of between course ratings. "Because the focus of the ratings is on overall teaching effectiveness, one should not be surprised to find a halo effect. The appropriate question is as follows: Does the halo effect invalidate students' overall ratings of teaching effectiveness? It probably does to the degree that concern for students' learning and other positive teacher characteristics are overweighted by students in their overall judgment" (p. 1221). Therefore, students who feel the instructor

does not care about their learning develop a negative halo while just the opposite occurs for those who believe the instructor cares. Balzer and Sulsky (1992) provide a critical examination of halo and performance appraisal research, pointing out a number of perceived flaws including lack of consensus for a conceptual definition of halo, a lack a systematic relationship between conceptual and operational definitions of halo, halo measures perhaps being poor indexes of rating quality, and even that halo may or may not be an important measure of the rating process. In yet another review of halo research, Murphy, Jako, and Anhalt (1993) question whether halo effect is even necessarily bad. They cite studies that indicate that halo error does not automatically limit rating accuracy and may actually increase accuracy in some circumstances.

Whereas Balzer and Sulsky (1992) conclude that halo may be more appropriately used as a measure in studying cognitive processing rather than as a measure of performance rating outcome, Murphy et al. (1993) actually call for a moratorium on using halo indices as criteria for evaluating training programs, rating scales, and so on. Their review addresses seven common beliefs about halo error, which they demonstrate are not necessarily true. These beliefs are that halo error is thought to be common; that halo error reflects the rater's inability or unwillingness to discriminate among different attributes of the individual rated; that observed halo (the average intercorrelation among ratings) reflects the influences of true halo and illusory halo; that halo error causes inflated correlations among the rated attributes; that the rater's overall evaluation of the rate influences the ratings of specific attributes; that halo error has a negative impact on the quality of ratings; and that removing halo error is both plausible and worthwhile. Instead, they call for redirecting halo effect research to study whether and when a halo error has occurred, how contextual factors affect measures of

observed halo, and when halo errors are harmful or beneficial. Additionally, Becker and Cardy (1986) assert that “determination of the usefulness of halo as an indicant of accuracy remains a difficult task, even when estimates of true performance levels are available” (p. 669). They call for consideration of alternative measures of halo and accuracy along with the issue of relations due to construction of the measures.

Moreover, Solomonson and Lance (1997) find that individual differences in thought processes that raters use may not be as widespread as has been thought. Their study determines that true halo (the correlations among the actual behaviors being rated) and halo rater error (influence of a rater’s general impression) taken together account for a substantial proportion of observed halo, and the findings support a null true halo-halo rater error relationship. In other words, little difference exists in the level of halo rater error across varying levels of true halo. This means that to the extent that raters rely on general impressions of those they rate because of processes that categorize, the raters’ general impressions are relatively homogeneous across rating dimensions. A study that Nathan and Lord (1983) conducted comes to similar conclusions. They assert that statistical approaches to controlling rating error will conclude in “equivocal success” in reducing halo. They call for more study to understand how people process information.

As Becker and Cardy (1986) also point out, those who study halo error generally use a variance measure, reflecting the extent of intraratee ratings differences across performance dimensions, or a correlation measure, examining the interdimensional covariance of each rater’s appraisal. They cite research that finds these two measures to be virtually independent, and their own study concludes that variance and correlational forms of halo and

accuracy can yield identical, corresponding, or widely divergent results as well as that results appearing to be counterintuitive can often be explained as a necessary outcome given the construction of the halo and accuracy measures. Finally, they contend that controlling for measured halo eliminates as much true halo as error halo, thus making the process not useful.

Although not discussed in terms of halo effect, psychological research concerning attitude structures also appears to relate to both halo effect and the potential relevance for student ratings of teaching. To delve extensively into this literature is beyond the scope of the current study; however, it has implications that bear investigating. For instance, Judd, Drake, Downing, and Krosnick (1991) contend that attitudes are object evaluations stored in memory and that giving one attitude response could facilitate a second attitude response if the second shares a structural link in long-term memory. Their study of attitude structures finds that spreading activation can polarize attitude responses so that answering questions on one attitude might result in more extreme responses to a second linked attitude issue. In relation to SRTs, such results indicate that certain questions on an evaluation form might trigger long-term memories of events that cause attitudes to emerge that, in turn, could affect other responses to related topics. One might theorize on the basis of these findings that students whose experience has been positive might tend to provide more positive ratings whereas just the opposite could occur with students who have developed a negative attitude about a particular class or instructor. The relation to halo effect is clear, and these results appear to relate to the findings of Lance, LaPointe, and Stewart (1994), discussed above, that indicate that halo error results for ratees forming a general impression, which then colors all the ratings. Whether a particular incident or set of incidents in a course creates an attitude about the course or instructor that is powerful enough to skew ratings slightly in either a more

positive or a more negative direction constitutes a hypothesis that deserves additional study.

Kishor (1995) provides yet another explanation for the occurrence of halo effect. In his study of the operation of implicit personality theories in relation to biases in performance ratings, his findings show that behavioral information can imply traits and traits imply unavailable behaviors. He asserts that students infer unavailable traits and behaviors from given behaviors because traits and behaviors are encoded together. Therefore, the inferred ratings most likely measure covariance of traits and behaviors in the raters' implicit personality theories of teaching rather than in the instructor's teaching behaviors. Additionally, he claims that the findings imply that inferential bias can still affect performance rating instruments free of trait items. He concludes that the mental comparisons and inferences therefore automatically bias the ratings, which is why the halo error is so ubiquitous in performance ratings.

Overall, the various research studies indicate that halo effect is nebulous and may or may not constitute a serious factor creating bias in student ratings of teaching. Moreover, with some reviews questioning the reliability of halo measures because of lack of agreement on common conceptual and operational definitions of halo and whether it is even an important measure of the rating process, they cast doubt on how far one can or should even consider halo effect in relation to student ratings of teaching.

Discriminant Validity (Bias) Studies Related to Instructor Variables

As discussed earlier, discriminant validity studies examine what Freedman, Stumpf, and Aguanno (1979) call "extrinsic" variables that fall outside the realm of instructor control. Characteristics such as gender, rank, class size, and so forth, which could influence how

students rate courses or instructors, would represent bias if, indeed, they actually influence the results. Researchers have studied a variety of potentially biasing factors that might fall into the category of relating to the instructor or being under the instructor's control. These factors include age, gender, teaching experience, race, personality, research productivity, rank, and status (fulltime versus adjunct). The instructor variables that have received much of the attention in the literature deal with grade leniency, instructor likeability (rapport), and instructor enthusiasm (the Dr. Fox effect) and their possible influence on student ratings.

Non-Bias Factors.

Variables related to the instructor that have little or no biasing effects include age, gender, teaching experience, race, personality, or research productivity (Cashin, 1995; Feldman, 1986, 1993, 1997). However, El-Hassan (1995) finds that students give higher ratings to more experienced instructors, meaning instructors who had more years of teaching experience, but finds a lack of any relationship between mean ratings and instructor rank, degree, and status (part-time vs. full-time). Based on a review of studies, Marsh (1987) also concludes that measures of research productivity have no bearing on teaching effectiveness or vice versa.

Rank.

One could consider rank as a variable that may bias student perceptions of faculty in instances in which students may associate higher rank with greater teaching effectiveness, which, in turn, causes students to have a preconceived evaluation before ever experiencing the class if that preconception continues to color their opinions throughout the course. The findings in relation to this factor are somewhat mixed but generally indicate that the effect of

rank on student ratings may be slight at most, yet Aleamoni (1987) cites studies showing mixed results and contends that rank as a factor that affects student ratings is a myth.

Marsh (1980) reports SEEQ research results indicating that teaching assistants receive lower ratings than regular faculty for most rating dimensions and overall rating items. However, teaching assistants may receive slightly higher ratings for the individual rapport and group interaction dimensions. But excluding teaching assistant ratings, Marsh discovers almost no relation between rank and global ratings but a somewhat positive correlation with the breadth of coverage dimension and a somewhat negative correlation with the group interaction dimension. Feldman (1983) generally supports these findings, asserting that positive relations are more likely for dimensions related to instructor knowledge and intellectual expansiveness. Ratings related to encouragement of discussion, openness, and concern for students, on the other hand, reveal negative relations.

Wigington, Tollefson, and Rodriguez (1989) examined instructor rank, along with reputation and sex, and claim that no relationship occurs between the rank of an instructor in relation to the type of class taught (lecture, discussion, lecture-discussion, or laboratory) and little relationship between rank and level of the course (lower, upper, or graduate). However, the highest mean rating appears in connection with the relationship between reputation and rank, with full professors for whom students report that reputation is a factor in their decision to take the course receiving the highest ratings. Conversely, full professors whose reputation is not a factor in course choice earn the lowest mean rating. The smallest difference between mean ratings appears for graduate teaching assistants. These results indicate that over time, faculty who earn positive reputations will likely receive higher student ratings, but Wigington et al. contend that such results suggest that professors developing positive reputations as

effective teachers also receive higher student ratings. Furthermore, they speculate that these consistently higher ratings reflect well-developed teaching skills.

Instead of considering strictly the rank of the instructor, some studies translate rank into years of experience.

Years of Experience.

The contention involved with years of experience as a potential bias is that ratings may be increase based on years of experience. The research generally finds either no relation or a negative relation between years of experience and student ratings.

An older review of literature by Costin, Greenough, and Menges (1971) lists studies dating back to 1936, which find either no significant relationship or that some student ratings improve with the experience of the faculty member. It stands to reason that experience should lead to greater knowledge of subject matter, more effective presentation of subject matter, and other factors that could result in higher student ratings, but it is unclear from what Costin et al. report as to whether experience (or rank) per se is a biasing factor or whether experience translates into more effective teaching, which would not, therefore, be a biasing factor. Kulik and McKeachie (1975) cite additional older studies that found no significant relation between teaching experience and student ratings.

In his review of other studies, Feldman (1983) noted that about half the studies discover no relation between either age or years of teaching experience and global ratings. On the other hand, among those that find significant relations, the predominant result is a negative relation.

Experience related to having taught a course more than once shows a positive correlation with higher student ratings, according to Marsh (1982a). Marsh asserts that when

the same instructor teaches the same course in different terms, the ratings tend to be higher the second time. Such a finding indicates teachers improve with practice and thereby signals that higher ratings related to years of experience may result from simply being better instructors. In this event, years of experience would not constitute a bias.

Type of Instruction.

In their study looking for correlations between three class variables (type, level, and size) and three instructor variables (reputation, rank, and sex), Wigington, Tollefson, and Rodriguez (1989) find the second strongest effects coming from class type. They examine classes taught in one of four formats: lecture, combination lecture and discussion, discussion, and laboratory. Results indicate that, overall, instructors leading discussion classes have a higher mean rating than instructors employing the other formats. Instructors using the lecture-discussion format receive lower ratings when the class size is small (25 or fewer students) than when it is large (50+ students), and instructors using the laboratory format receive lower ratings for small and large classes than for moderate-sized (26-49 students) classes. They posit that class type is a powerful factor in differentiating faculty who receive higher and lower student ratings. Reasoning that students' preference for a discussion format likely indicates the well-established importance of active student involvement in learning, they assert that an implementation of an active teaching-learning process, such as discussion, could lead to higher ratings. However, to the extent that students are reacting to the type of instruction in relation to more effective learning, this would not qualify as a bias. If students choose classes on the basis of the format of the class as a personal preference or for any other reason than as their preferred method of more effective learning and if students who prefer discussion classes are also predisposed to rating more highly those instructors who use this

type of delivery, then course type would, indeed, constitute a bias.

Grading Leniency/Expected Grade.

The factor that appears to have received the most attention without being resolved concerns grading leniency. As Gigliotti and Buchtel (1990) point out, “Although it is intuitively appealing to believe that higher evaluations result when students receive high grades or higher grades than expected, the research on the issue is unclear” (p. 342). Little evidence in their study supports the contention that either easy or hard grading affect ratings of instructors. Furthermore, Aleamoni (1987) lists grading leniency affecting student ratings as one of the myths that faculty members hold, a view he bases on the results of dozens of studies.

Howard and Maxwell (1980) present the results of their path-analytic study of three models designed to investigate the correlation between student satisfaction and grades: a grading leniency bias model, a teaching effectiveness model, and a student characteristics model. They conclude that, despite some amount of influence of grades on satisfaction, the influence is less than prior literature suggests. Rather than grading leniency as causing the relationship between grades and student satisfaction, they conclude that student motivation plays a more important role in accounting for the relationship.

Determining whether grading leniency or an expected grade reflects a measure of teaching effectiveness or a true bias confounds this potential bias (Marsh, 1982a). Howard, Conway, and Maxwell (1985) put the dilemma in slightly different terms. “Finding a relation between grades and student evaluations might equally well suggest the influence of grades on student evaluations and/or at the same time represent evidence for the validity of student

evaluations” (pp. 187-188). Although Marsh (1983) found that a higher expected grade was one of four factors most predictable of student rating variances in the SEEQ, he says little to explain how it matches with a specific dimension in the SEEQ but claims that an interpretation of higher expected grades as being a bias to student ratings “seems unwarranted” (p. 165). Yet in his study of the same instructors teaching the same courses at different times, he concludes that the results favor the explanation that more effective teaching causes better course mastery and, hence, better student ratings rather than their being a grading leniency bias. As an alternative explanation for positive results in multisection validity studies, Marsh (1984), in a subsequent study, posits what he calls the grading satisfaction hypothesis, or what others label as the grading leniency effect. When a significant correlation exists between the known or expected course grades and performance on the final exam, then higher student ratings may be due to more effective teaching that, in turn, results in greater learning and higher ratings; due to increased student satisfaction with higher grades that causes students to rate the instructor higher, independent of more effective teaching or greater learning, as a “reward”; or due to initial differences in student characteristics, such as prior subject interest, motivation, or ability, that affect both teaching effectiveness and performance. According to Marsh, the first hypothesis affirms the validity of student ratings as a measure of teaching effectiveness, the second constitutes a bias in the ratings, and the third results from presage variables that student ratings may accurately reflect. The major debate concerns whether grading leniency is a bias.

An earlier study that Stumpf and Freedman (1979) conducted presents evidence that expected grades correlate positively with instructor ratings at both the individual and the class levels of analysis, with more instructor rating variance occurring at the class level than the

individual level. They explain this difference as perhaps resulting from instructors' grading policies having a greater influence on classes of students than their individual expected grades but that it is also possible that the differences in class expected grades reflect differences in the amount students in various classes learn. In essence, expected grades may or may not constitute a bias. They conclude by asserting that their findings suggest that researchers should consider other possible sources of variance in the rating system such as halo effect besides traditional response biases.

A more recent study by Salmons (1993) indicates that student expectations of grades influence SRTs. According to the study results, students expecting to fail significantly lower their ratings whereas student expecting to receive an "A" or "B" raise their evaluations significantly. However, Salmons also acknowledges that student grades account for only a small percentage of the evaluation variance. Furthermore, Salmons names other factors such as lecture style, ability, and willingness to provide assistance outside the classroom as possibly attributable to students' evaluations. Even by factoring out expected grade, ineffective instructors will still receive low ratings and excellent instructors will still receive high ones, indicating that even if grade expectation is a factor, it is not a significant one.

In their study of attributional bias in relation to student ratings, Gigliotti and Buchtel (1990) explore both the self-esteem model and the expectancy-confirmation model. The self-esteem model relates to the idea that people tend to attribute their successes to their own efforts but blame their failures on others. With grades providing a measure of success or failure relative to the students' expectations, the self-esteem (or self-serving) model predicts that students will give instructors lower than deserved ratings because they will attribute a

high grade to their own effort but a low grade to some failure on the instructor's part. The expectancy-confirmation model contends that current motivation and past experience with an achievement context provides the basis for students' expectations. Thus, they will attribute the successful meeting of outcomes to self and failure to the instructor. Gigliotti and Buchtel point out that this second model predicts both a grading-leniency and a grading-stringency effect. In other words, students with low expectations who perform poorly and students with high expectations who perform well will attribute their failure or success to self; students who expect to do poorly but receive a higher grade and those who expect a higher grade but perform poorly will attribute their performance externally to the instructor. In fact, the results of their study show that self-esteem bias has a minimal or nonexistent effect on evaluations based on an analysis of both individual and class means. They find that although low grades are associated with negative instructor evaluations (external), as the self-esteem model predicts, low grades are also associated with negative attributions to the self (internal), which the model does not predict. Furthermore, the data reported under the expectancy-confirmation model reveal that despite evidence that both effects may have been operating, neither was very strong. Grading leniency bias will be occurring for the low students who receive high grades and the high students who receive high grades, thus resulting in higher ratings according to the model; however, the results show that although higher ratings do occur for fairness, they do not occur in conjunction with other factors: skills, support, or comparative rating. In terms of the grading-stringency effect, one can check for bias by comparing the two low outcome conditions (low/high expectations with low outcomes). According to this expectation, lower evaluations should come from those students with high expectations but low outcomes; however, the data fails to confirm this bias. They conclude

by saying that their results support other results that grades cannot be ruled out as a potential biasing factor but that the relationship between actual grades and evaluations is most apparent at the individual level and not at the class level; that expected grade and expectancy violation are ruled out as biasing factors; and that how actual grades relate to ratings hinges on the level of analysis and the evaluation dimension being studied. They also contend that the self-esteem or self-serving bias effect documented in attribution literature only minimally affects instructor ratings. Costin, Greenough, and Menges (1971) report similar findings in their review of much earlier work.

In a 1972 study, Holmes finds that students who receive grades lower than their expectations tend to rate the instructor and course lower than those students who receive the grades they expected. These results offer a different implication than those studies that assert a grading leniency bias because in Holmes's study, that students might have received grades other than A's is not the factor that affects ratings. In effect, students are likely retaliating or displacing aggression in response to the lower grades. Since the lower ratings appear only with students who receive grades lower than expected, Holmes recommends that instructors should adequately inform students of their proficiency to decrease or eliminate any bias due to disconfirmed expectancies.

Some studies indicate that a bias exists, thus causing student ratings to be invalid. According to Powell's study (1977), as grading criteria increases, student ratings of instructor performance and course evaluation decreases. He finds that grades also strongly influence the nature of the comments that students write on the course evaluations, with the more stringently graded course receiving markedly more negative student comments. Worthington

and Wong (1979) conclude that grades modify evaluation ratings of instructors, with those with least mastery of the content more likely to give more positive evaluations when their grades are inflated. In effect, this finding implies that instructors can raise their student ratings by giving higher grades to students who have learned less. They suggest that the validity of student evaluations of instructors must be questioned seriously. Chacko (1982) studies two sections of the same course taught by the same instructor, covering the same content, using the same course materials, and giving identical exams. Students in the class were randomly assigned and were both relatively equal in test results. The only difference is that for one of the sections, the instructor graded one section harder than the other. The results indicate that students who are graded more harshly gave the instructor a much lower evaluation, leading Chacko to question “the use of student ratings as a means of obtaining unbiased information regarding teaching effectiveness” (p. 24). In effect, Chacko contends that student ratings reflect subjective bias on the part of the students, at least in part determined by their grades.

Greenwald and Gillmore (1997a), in a study at the University of Washington, determine that grading leniency does have an effect on student ratings. Such a finding indicates that discriminant validity in student ratings fails despite the repeated studies demonstrating moderate convergent validity.

The results of a number of other studies fail to discover a grading leniency bias. Howard and Maxwell (1980, 1982) report that by controlling for prior student motivation and student progress ratings, they can eliminate most of the covariation between expected grades and class-average overall ratings. They conclude that student motivation influencing performance, grades, and satisfaction is a more potent contributor to the covariation between

grades and satisfaction than is the direct contaminating effect of grades upon student satisfaction. In another set of controlled laboratory studies involving grading leniency, Abrami, Dickens, Perry, and Leventhal (1980) present significant ($p < .05$) effects of grading standards related to student ratings, but these effects are inconsistent for all types of ratings, instructors, or standard differences. Therefore, they conclude that grading standards fail to affect student achievement in either experiment and that the differences are relatively unimportant when using ratings to make gross distinctions between instructors. Frey, Leonard, and Beatty (1975) find no significant correlation between grading and student performance on the final exam. The results of a meta-analysis by Feldman (1976a) concludes that although the available evidence cannot definitely establish a bias in instructor evaluations due to the grades students receive or expect to receive, the evidence also does not preclude such a bias from existing.

Greenwald and Gillmore (1997b) classify the positions studies and reviews take into three categories: those that consider grading leniency to bias student ratings of effective instruction (i.e., SRTs are invalid), those that take an intermediate position between these two extremes (i.e., SRTs may be invalid but are still useful), and those that contend grading leniency presents only a minor biasing factor at most (i.e., SRTs are valid). They seek to evaluate theories of causal influences affecting student ratings. Various theories of these influences imply different patterns of relationships among three categories of measures: evaluative ratings, expected grades, and course workloads. Their study covers three consecutive semesters of student ratings for several hundred courses at the University of Washington. Of the various theoretical models proposed, the Grading Leniency model implicates a strong influence of grading policy on student ratings. They had hypothesized

that grading leniency differences lead directly to differences in expected grades (lenient grading on midterm exams should cause students to expect higher final grades) and that since leniency-caused high expected grades should exceed students' grade aspirations, a positive correlation between expected grades and evaluation ratings should occur, resulting in grade satisfaction effects. They find student ratings to be sensitive to grading leniency, a factor that SRTs are not intended to measure. In a separate article, Greenwald and Gillmore (1997a) contend, however, that these results should not cause the abandonment of SRTs. Despite what they see as a deficiency in discriminant validity, users cannot and should not dismiss the evidence for convergent validity of SRTs.

In a more recent article, Gillmore and Greenwald (1999) present a rationale for their findings of grading leniency, considering it "an unintended side-effect of conscientious efforts to improve their teaching" (p. 519). They contend that strict-grading instructors likely encounter student suggestions to make changes to improve teaching with these suggestions being related to the assignment of too much reading or devoting insufficient class time to difficult material. Consequently, instructors revise their syllabi to reduce the amount of material covered, enabling better course mastery, leading to higher grades and, ultimately, higher ratings. Although they acknowledge that a few instructors may intentionally give students higher grades in an attempt to increase their student ratings, the above scenario involving indirect leniency is more likely what occurs.

"The small, positive grade-SET [Student Evaluation of Teaching] relations are probably the most hotly debated topic in the literature on potential SET biases" (Marsh & Roche, 2000). Marsh and Roche (2000) present three main hypotheses interpreting the issue of grade relations: the validity hypothesis (higher grades reflect better student learning and that a

positive correlation between student learning and SRTs supports the validity of SRTs); the prior characteristics hypothesis (preexisting student or course variables such as prior subject interest, prior motivation, class size, or course level affect student learning, grades, and actual teaching effectiveness such that the grade relation may be spurious); and the grading leniency hypothesis (instructors who give higher than deserved grades are rewarded with higher than deserved SRTs, constituting a serious bias to SRTs). In addition to these three, Greenwald and Gillmore (1997a) include an attribution theory similar to the one that Gigliotti and Buchtel (1990) present as discussed earlier.

Right behind grading leniency in the list of factors that faculty members generally assume is a bias, according to Marsh (1982a), is instructor popularity. Instructor personality, enthusiasm, and likeability (rapport) all may be contributing components of an instructor's popularity.

Instructor Personality.

A common belief among some faculty members is that it is the "popular" instructors who wins high ratings because of their personality traits such as being warm and friendly or displaying a sense of humor in the classroom. The issue of instructor personality also relates to other potential biases connected with instructors, including enthusiasm (or the "Dr. Fox effect") and likeability (or rapport), discussed below. In their review of older literature, Costin, Greenough, and Menges (1971) cite a few studies that find personal characteristics such as having a "pleasing personality"; being "sympathetic," "helpful," "sincere," and "enthusiastic"; having an "interest in students"; and treating students with "fairness" and having a sense of "humor" are all ones that students list as desirable. Kulik and McKeachie (1975) also mention some studies related to a teacher's "style" as opposed to "substance" as

well as studies indicating that instructors with good communication and presentation skills receive higher ratings, the implication being that these personality traits can affect favorable student ratings. All such characteristics may, in some way, contribute to the overall likeability of an instructor. However, based on their review, Costin et al. (1971) conclude that little evidence exists to demonstrate whether personality impacts student ratings of instructors' performance. Although Feldman (1990) in his meta-analysis finds that instructor's personality, friendliness, and respect or concern for students all correlate with student achievement, these are less strongly correlated factors than 12 other instructional dimensions, and in the final analysis the dimension of instructor personality is not included as one of the 6 dimensions that explain at least 10% of the variance in student achievement: preparation and organization, clarity and understandableness, stimulation of interest, high standards and motivation of students, encouragement of discussion and openness to opinions of others, and elocutionary skills.

Marsh (1987) asserts that instructor personality is important for at least two reasons. First, some suspect that personality has a substantial effect on ratings but has nothing to do with being an effective teacher. If that is the case, then instructor personality will be a potentially biasing factor in students' ratings. Second, a significant relationship will indicate that the results may have practical and theoretical importance for distinguishing between effective and ineffective teachers as well as for better understanding of the study of teaching effectiveness and the study of personality.

Relatively few studies appear examining the relationship of personality to ratings. Feldman (1986) reviews studies relating student ratings and 14 categories of personality inferred from instructor self-reports and ratings by students or colleagues. The only positive

correlations occur with positive self-regard and self-esteem and with energy and enthusiasm. The instructor self-reports systematically relate to students' overall evaluations of the instructors, but the correlations of reports by students or colleagues are substantially larger than those based on self-reports. The results of this review indicate a relation between student ratings and instructor personality, but it is unclear whether these results indicate a valid source of influence or a source of bias. In fact, according to Feldman, little appears in the research he analyzes to support contentions that college instructors' effectiveness as perceived by students has any association with personality characteristics. According to Marsh (1987), the problem with the interpretation based on existing research is that the research confounds the variables of self-report versus report of others (students or colleagues) and the variables of general versus situationally specific aspects of personality (personality as an academic professional versus personality as a teacher). However, Aleamoni (1987) contends that it is a myth to believe that student ratings are nothing more than popularity contests. He cites his own earlier work and other studies indicating that students can praise an instructor's warmth, friendliness, and humor in the classroom yet still criticize the same instructor for being disorganized or for using teaching methods not conducive to student learning.

Enthusiasm (the "Dr. Fox Effect").

A key characteristic of instructor personality and the one that some may believe is the main contributor to instructor popularity is enthusiasm in the classroom. Abrami (1989b) asserts that the Dr. Fox studies are the most often cited evidence against the validity of student ratings. Early studies focus more on the effects of instructor enthusiasm on student learning, but Ware and Williams (1975) also include the impact of enthusiasm on student

ratings as well as student learning of academic content. This early study reports that when an instructor's presentation is highly expressive, student ratings of instruction are just as high regardless of any differences in course content or students' actual test performance. Williams and Ware (1977) reconfirm these findings in a follow-up study, concluding that student ratings of high-expressive instructors may not reflect how substantial the course content of their classes are nor the degree of student achievement. These findings, in particular, launched a series of other studies aimed at trying to confirm instructor enthusiasm as a biasing factor in student ratings of teaching. As Cruse (1987) asserts in referring to the Dr. Fox factor, students rate enthusiastic and informative performances justifiably but also rate the enthusiastic but less informative instructor just as highly. One of the reasons students may like certain instructors more than others comes from the enthusiasm the instructor conveys in the classroom.

However, subsequent studies have pointed out the methodological weaknesses with Ware and Williams's (1975) early studies, including a lack of an adequate control group, a poor rating instrument, and problems of generalizability to college classrooms (Abrami, Leventhal, & Perry, 1982; Marsh & Ware, 1982; Perry, Abrami, & Leventhal, 1979). In the study Perry, Abrami, & Leventhal (1979) conducted, they fail in their attempt to replicate the results of the research that Williams and Ware (1977) conducted and find that unlike the results from their studies in which content does not affect ratings and achievement similarly for highly expressive instructors, in the study by Perry et al. (1979), content does have an effect. Furthermore, Braskamp, Ory, and Pieper (1981) present their results that in analyzing students' written comments to open-ended questions about the instructor, students comment on the instructor's knowledge of the field and ability to communicate more frequently than

they mention conscientiousness or enthusiasm. Their findings raise a question concerning whether it is possible the instructor enthusiasm may relate to student perceptions of an instructor's communication skills and that perhaps enthusiasm becomes confounded with these skills.

“The Dr. Fox effect is defined as the overriding influence of instructors' expressiveness on SETs [Student Evaluations of Teaching] and has been interpreted to mean that enthusiastic lecturers can ‘seduce’ students into giving favorable evaluations, even though the lectures may be devoid of meaningful content” (Marsh & Roche, 1997, p. 1193). Marsh (1983), Marsh and Ware (1982), and Marsh and Roche (1997), however, contend that what may seem like a bias in some research is actually related to “Enthusiasm,” one of the dimensions of effective teaching Marsh uses in the SEEQ, and therefore serves to illustrate the importance of multidimensionality in SETs. The meta-analysis Abrami, Leventhal, and Perry (1982) conducted also supports this finding that expressiveness may have a greater relationship to individual factors related to enthusiasm rather than to factors not logically related to it. They assert that effect sizes reported in various studies are larger in individual factors logically related to expressiveness, such as instructor personality and instructor expressiveness, but smaller if a factor contains unrelated items, such as lecture content, instructor preparation, and instructor explanation.

Marsh and Ware (1982) assert that although instructor expressiveness produces better student ratings when no student incentive is involved, content coverage is more important. They argue that because grades act as the incentive in a normal classroom environment, expressiveness is likely not a biasing factor. They conclude that, “Apparently the use of humor, voice inflections, and enthusiasm are important when extrinsic motivation is low, but

are less important when students are already motivated” (p. 133). Moreover, it is possible that enthusiasm may be one of several factors that combine to affect student ratings of instructors. Enthusiasm for the subject matter and for teaching appears as one of the characteristics consistently associated with superior college teaching according to a meta-analysis of research studying student views of quality instructors and instruction (Feldman, 1976b). However, Feldman includes enthusiasm with other factors (instructor’s stimulation of interest, clarity or understandableness, knowledge of subject matter, and preparation and organization) as having a greater association with overall ratings of superior instructors or instruction. Therefore, it is unclear from Feldman’s meta-analysis whether these other factors along with enthusiasm combine to produce the higher ratings or even if enthusiasm is a bias or an instructor characteristic that produces greater student learning. Results from a later meta-analysis (Feldman, 1990) indicate that although enthusiasm is one dimension of instruction that correlates with student achievement, it is less strongly correlated than several other dimensions, including instructor preparation and organization, clarity and understandableness, instructor’s pursuit and/or meeting of course objectives, student-perceived outcome or impact of the course among others.

Although an enthusiastic delivery style likely does relate to a dimension such as Marsh’s Enthusiasm, results from a study by Williams and Ceci (1997) indicate that a factor such as enthusiasm may also create a halo effect, impacting the overall ratings of individual factors as well as the course in general. Basically, a halo effect occurs when students indiscriminately assign high ratings to all factors of a course or instructor based on an overall impression of the instructor rather than actually considering each factor independently. In their study, Ceci replicates a course from fall semester to spring semester with the only

change being in teaching style. Ceci received training in delivery, specifically varying voice pitch more and using more hand gestures. As expected, the factor of enthusiasm increased, in fact dramatically, from 2.14 (fall) to 4.21 (spring)—an increase of over two standard deviations ($t = 25.58, p < .0001$). However, more surprising are the findings that all factors of ratings of instructor (knowledgeable, tolerant, accessible, and organized) in addition to “enthusiastic” as well as all factors of the ratings of the course (how much learned, clear goals, fair grades, text quality, overall course rating, and recommend course) increase as well (p. 20). Williams and Ceci are also shocked to discover that the students in the spring semester are wrong about how much they had learned. Ratings indicate that they had learned more when, in fact, the objective tests and final grades given in the two semesters were virtually identical. Although they claim that their point is not to indicate that a halo effect has occurred, the results imply that change in one variable—instructor enthusiasm—impacts the ratings of all factors for the entire course. “Given all of the limitations of a naturalistic case study, our modest study nevertheless shows that student ratings are far from the bias-free indicators of instructor effectiveness that many have touted them to be” (p. 23).

In their meta-analysis, Abrami, Leventhal, and Perry (1982) find that in all studies, instructor expressiveness has a significant and reasonably large effect on global ratings, i.e., general questions about the course and the instructor. But the effect of content on student ratings for high expressiveness is almost the same as the effect of low expressiveness. In other words, expressiveness has relatively little impact on student learning, a finding that Feldman (1990) also confirms. According to their findings, instructor expressiveness has a substantial impact on summary and global student ratings but much less on student achievement whereas lecture content has a substantial effect on student achievement but

much less on student ratings. However, they also contend that because of poor comprehensiveness, Dr. Fox research “tells us little about the validity of ratings, regardless of the findings” (p. 457). Feldman (1990) also finds that both students and faculty regard enthusiasm as being of high relative importance in teaching, yet enthusiasm is only moderately important in association with students’ achievement and students’ overall evaluations of teachers. The implication of these results is that perhaps both students and faculty give enthusiasm more credit than it is due in terms of the effect it has on students, certainly on student achievement. The mixed results make it unclear whether enthusiasm actually constitutes a substantial bias in student ratings. Besides discussions of enthusiasm, various discussions of student ratings sometimes connect instructor enthusiasm with the factor of instructor likeability, indicating that perhaps a connection may also exist between instructor enthusiasm and likeability or rapport.

Instructor Likeability (Rapport).

Regardless of whether likeability constitutes a bias of student perceptions that students then transfer to SRTs, instructors often voice the opinion that SRTs are merely “popularity contests.” Cruse (1987) uses the term “charismatic personality,” which indicates a possible cause for this instructor likeability. Regardless of whether one labels the characteristic likeability or charismatic, the consequence is that some faculty members seek reasons for some instructors being more popular than others. Trout (1997), for example, mentions that instructors merely need to be funny to raise their evaluation scores. Although I discuss grading leniency and workload reduction as other factors suspected of biasing student ratings, both Trout (1997) and Martinson (2000) imply that some instructors resort to these tactics to raise their ratings, yet one might argue that both of these factors are simply symptoms of

other motives, such as wanting students to like them with the reward being higher ratings, attempting to reduce student anxiety, or simply trying to get student attention.

In a study of 100 students, Waters, Kemp, and Pucci (1988) asked students to consider the two or three instructors to whom they had given their highest and lowest evaluations. Then the students listed the aspects of the instructor or course that had most influenced their ratings. The study reports no difference between introductory and upper-level courses or between genders of the students. However, marked differences among the factors student consider when rating instructors high versus low occur.

Generally, three categories emerge: personal characteristics (personal or inferred motivational characteristics), class characteristics (what occurred in class), and interpersonal characteristics (interaction with one or more students outside of class). Those instructors receiving the highest student evaluations garner them for personal and motivational characteristics, such as “enthusiasm, personality, sense of humor, and enjoyment of teaching,” and for interpersonal characteristics, such as “helping students outside of class and learning their names.” Students do not blame the low-rated instructors for lacking these qualities. Instead, students rate them low in terms of classroom behavior, such as “discouraging class participation, coming to class late or not at all, and lecturing without variation in class routine” (Waters, Kemp, & Pucci, 1988, p. 204). In effect, the study claims they assign positive personal and interpersonal characteristics to the high-rated instructors while the worst instructors receive impersonal ratings through comments about the way the instructors manage the class. The implication is that students give high ratings to instructors they like, at least in part, because of the personal and interpersonal characteristics. The

authors conclude “that ‘good’ faculty may be evaluated differently from ‘bad’ faculty and that characteristics leading to high evaluations do not necessarily have bipolar opposites leading to low evaluations” (p. 204). These findings reinforce the need to investigate what students think when they are rating instructors and what perceptions cause them to rate an instructor high or low, particularly if the “bad” instructors are not the polar opposites of the “good” ones.

Delucchi (2000) arrives at similar conclusions, finding that instructor likeability, relative to student perceptions of learning, effective teaching behaviors, and student effort, “exerts a disproportionately strong effect on overall [student evaluations of teaching]” (p. 221). More specifically, he finds that likeability exerts a statistically significant negative effect on student perceptions of learning, i.e., “the more students like an instructor, the less learning they report” (p. 225). Additionally, likeability exerts a positive and statistically significant effect on overall ratings despite controlling for the students’ perception of learning, and likeability proves a significant positive predictor of overall ratings. Delucchi contends that instructors who emphasize likeability can produce higher overall ratings compared to those stressing effort and enhancing perceptions of learning. The results indicate student biases in ratings, particularly in global items of teaching effectiveness.

On the other hand, in a much earlier study, Doyle and Whitely (1974) find that students’ liking for the subject correlates with ratings of how much they learn, the motivational effectiveness of the instructor, and overall instructor effectiveness; however, students’ liking for the subject matter does not correlate to their liking of the instructor as a person or to their ratings of his or her general teaching ability or attitudes about teaching. Their findings cast doubt about whether likeability is truly a biasing factor.

As mixed as the results are for instructor-related factors, the ones that relate to students prove to be just as mixed, with studies presenting evidence both for and against various variables being biases. Additionally, the research investigating student-related variables is as extensive as the studies dealing with course- and instructor-related ones.

Summary of the Research

“Student evaluations are the most common, yet most controversial source of information on which to judge effective teaching....While there are many factors that substantially influence student ratings, it is safe to say that there is a reasonably positive relationships between student evaluations of teaching and academic achievement” (Simpson, 1995, p. 4). Beyond these two observations, it is difficult to claim much about student ratings with any certainty. As indicated in this chapter, research results offer evidence both supporting and negating the reliability and validity of student ratings of teaching. As Simpson also points out, concerning student ratings, some contend that the classroom is the faculty’s domain, that teaching effectiveness is impossible to identify and measure, and that students are incapable of evaluating in a valid manner the quality of instruction they receive and should not have the opportunity to do so. On the other hand, others believe that instruction is measurable and that student ratings are both reliable and valid and that students are capable of rating all instructors. What one study might assert as a bias, another study will contend is simply a measure of effective instruction. Clearly the research related to student ratings of teaching demonstrates the complexity of the topic as well as the fact that studies have arrived at few definite conclusions. Stumpf and Freedman (1979) indicate the complexity in the closing of their discussion about their study of grading effects when they

declare that a combination of the situation, the ratee, the rater, and the instrument in concert and interaction create variance in ratings and that measurement error, bias, and validity are all possible interpretations of portions of this variance. “A more general approach to investigating criterion variance is necessary if diverse research findings on evaluation issues are to be understood and integrated into meaningful prescriptions” (p. 300). As the review of the studies in this chapter illustrates, this integration has yet to emerge, and little can be said with any certainty about SRTs. Most researchers appear to agree that rating instruments are generally reliable and achieve generalizability; however, far less agreement occurs in discussions of validity. Despite general agreement about common dimensions of effective teaching, far less agreement appears in terms of factors that may bias student ratings. Ultimately, the interrelationship among the four elements that Stumpf and Freedman name may all impact ratings to some degree and may ultimately be impossible to unravel.

Limits of Studies of Student Bias and Identified Needs for Further Research

Despite thousands of studies related to the reliability, validity, and generalizability of student ratings of teaching, the topic of student integrity has yet to appear in any of literature. As discussed in this chapter, research related to SRTs falls into four general areas: consequential validity, conceptual structure, convergent validity, and discriminant validity. Because the issue of whether students are candid and sincere in completing ratings could potentially bias the results of these ratings, it constitutes an aspect of discriminant validity. Why this subject has not been investigated appears puzzling but perhaps understandable for a number of possible reasons.

First, quantitative research related to SRTs predominates in the literature. All the studies referenced in this chapter, for example, are quantitative in design. If students choose

not to be candid or sincere in their ratings of instructors and courses, the present studies are not designed to measure those factors. Therefore, researchers may be more concerned with variables that are more specific and more easily measured.

Most studies focus on one or a few potentially biasing factors, such as instructor enthusiasm or the combination of grade leniency and workload. At best, when significance occurs, the researchers conclude cautiously, indicating that a factor has a slightly significant effect and may, therefore, be a potential bias if they have found any significance at all. In other words, researchers enter a quantitative study isolating what they believe may be a potential bias or biases to the exclusion of other potential biases. The current study uses a qualitative approach to investigate the reasons students choose not to be candid or sincere, therefore allowing any potential biases to emerge.

More importantly, whereas current quantitative studies stop with making the point that a factor is or is not a factor that biases SRTs, perhaps because a quantitative design by its nature tries to isolate variables, this study employs a qualitative study to transcend knowing what the potential biases are to knowing some of the reasons why students choose not to be candid or sincere—to know more about what students are thinking when they complete SRTs.

As discussed in Chapter 1, recent research, points out that little is known about what students consider when completing SRTs (Wigington et al. 1989; Williams & Ceci, 1997) and has specifically called for more studies that investigate what students think when they are completing SRTs (Delucchi, 2000; McKeachie, 1997). To understand better what students think when completing SRTs appears better suited to a qualitative design. Given the scarcity of qualitative studies in the area of student ratings and no apparent studies dealing

specifically with student integrity while completing SRTs, this study will add new information which may assist faculty and administrators to make whatever modifications may be necessary to promote student integrity when completing SRTs and thereby help to improve teaching and learning.

CHAPTER 3

METHOD

Despite the extensive number of studies related to student ratings of teaching, they generally focus on the reliability, generalizability, and validity of the instruments themselves. As described in Chapter 2, these studies employ quantitative research methods primarily, and studies related to discriminant validity (biases) explore the possibility of various biases influencing the ratings themselves and try to account and control for these biases either in the instrument design or through statistical manipulations of the results. The attention remains more with the instrument itself than with the students. Calls for qualitative studies, as already mentioned, have been appearing within recent years in the literature (Delucchi, 2000; Greenwald & Gillmore, 1997a; McKeachie, 1997; Phillips, 1998; Williams & Ceci, 1997).

Research Design

Besides personal experience, indications both from the literature and from a pilot quantitative study that I had conducted provided indications that some students lack candor and sincerity in completing SRTs. Therefore, an assumption that I held entering the study was that some students choose to withhold what they really think and/or intentionally over- or underrate instructors, an assumption borne out by my study as the results in chapter four reveal. What was unclear before my study was why. A quantitative design, by itself, might have served to identify potential factors that may influence any lack of candor and/or sincerity on the part of students when completing ratings of teaching. However, a

quantitative study alone would not allow me to probe students for answers to that question why.

I had considered using a mixed design, using a quantitative survey to try to isolate potential causes, followed by a qualitative study employing interviews to probe for the reasons underlying those causes. However, such an approach would have likely overcomplicated my study, plus the literature is replete with quantitative data. Therefore, I decided to limit my study to a qualitative approach, specifically using a grounded theory approach (Strauss & Corbin, 1998). This approach enabled me to probe students' reasons for refusing to be candid and/or for choosing to be insincere when completing SRTs, and from this investigation my grounded theory about the core factor affecting student integrity emerged.

According to Strauss and Corbin (1998), a grounded theory study is “inductively derived from the study of the phenomenon it represents” (p. 23). My study analyzes the interrelated phenomena of student candor and student sincerity when completing student ratings of teaching. The nature of the phenomena lends itself to a grounded theory study because this design relies on using an inductive approach of constant comparative analysis to arrive at a theory concerning their ultimate cause since I had no definite conclusions entering the study and preconceived a variety of possible reasons. Therefore, an inductive approach permitted me to base my theory on what emerged from the data rather than beginning my study with a theory that I was attempting to prove. I chose a grounded theory approach as opposed to a phenomenological approach because I had a greater interest in determining a central underlying cause for a lack of candor and/or sincerity than in investigating the meaning of the phenomena. An interaction that influences the process of student ratings of

teaching occurs between instructors and students, and I wanted to arrive at a theory to explain that. Furthermore, by understanding the ultimate cause, both faculty and administrators have a better chance of acting in ways that will promote candor and sincerity rather than discourage it.

Participants and Site

The theoretical or target population entering the study was all community college students (and potentially all students in higher education institutions). The issue of student candor and sincerity when completing SRTs and the underlying motivations causing these behaviors apply to all students everywhere. However, for this study, the accessible population included only students at one community college in the western United States. This is an urban community college of approximately 7,500 students in a multi-college district. The student body is quite diverse ethnically. The only delimiting factor that applied to the selected sample of students relates to whether they had completed at least one student rating of teaching.

I was not part of the leadership of this college, and none of the students at this college would have any occasion to know me. Over the course of a year, I conducted interviews with 12 students. All the participants volunteered for my study resulting from three sources. One resulted from a presentation I made to the student government at the community college that served as the site from which my sample came. The purpose of this presentation was to inform them of what my study consisted of and to obtain permission to post a flyer to invite participation. Despite not specifically inviting any of the student government to participate, one of them contacted me later to volunteer. The second source of volunteers came as a result of the flyer that I posted around campus. However, I found this method to be generally

ineffective in reaching students as my flyer eventually became lost in a sea of flyers that apparently few students bothered to read or to which they simply failed to respond.

Ultimately, the majority of my volunteers came as the result of two faculty members who passed out my flyer in their classes.

Because all the participants volunteered, I had no ability due to human subjects restrictions to conduct any purposeful sampling to try to achieve maximum variation in demographics. However, ultimately, student demographics were not the important factors in this study. The only real prerequisite I had was that volunteers had to have completed at least one student rating. Even though demographics were basically irrelevant for my study, I conducted a demographic survey of all my participants out of curiosity just in case I discovered a demographic pattern of significance that would warrant focusing on a particular population. I found no patterns emerged.

Eight of the 12 participants were female. The participants' ages ranged from 19 to 54, with the mean age being 31, the median being 28. In terms of race, eight were Caucasian, one African-American, one Asian-American, and two of mixed heritage. Nine were fulltime students with all but two being in at least their third semester in college, the two exceptions being in their first and second semesters. Additionally, nine had completed more than one SRT. Only two of these participants have grade point averages below 3.0 with 2.58 being the lowest; 7 had grade point averages of 3.5 or higher with 4 of them having 4.0 averages. I discovered that common themes emerged from all 12 with only minor variations occurring and certainly no variations attributable to gender, age, or race differences. The mean grade point average for this sample was 3.5, the median being 3.61. What this high overall grade point average for this sample says is difficult to conclude with any certainty without further

investigation, but I speculate that these are not students who have any reasons to disparage instructors or have reasons to be less than forthright during the interviews. Additionally, because common themes emerged from the entire group, it seems reasonable to conclude that factors such as gender, race, age, year in college, and grade point average have little or no bearing on factors that affect their candor and sincerity when rating an instructor or a course. However, a correlation may exist between grade point average and choosing to volunteer for the study. If lower grade point averages are the result of lack of effort, for example, it is possible that these students also would not want to make the effort to participate in the study. Yet it may also be that students with lower grade point averages may have fears about participating in a study concerning student evaluations of teaching. Without specific investigation about this factor, it is impossible to draw any definite conclusions. For my participants, however, grade point average had no apparent influence on how they rate instructors. Table 1 summarizes the demographic characteristics of my participants.

Table 1
Summary of Participant Demographics

Gender	Age	Ethnicity	Status	Semester	Eval. Exp.	GPA
Female	24	Afr-Am	Fulltime	6	4+	2.58
Male	54	Afr-Am	Fulltime	6	2-3	2.91
Female	19	Caucasian	Fulltime	3	2-3	3.00
Female	28	Other	Part-time	7	1	3.20
Male	19	Caucasian	Fulltime	1	1	3.36
Male	26	Caucasian	Fulltime	4	4+	3.56
Male	20	Asian	Fulltime	2	2-3	3.66
Female	32	Caucasian	Fulltime	4	2-3	3.83
Female	28	Caucasian	Fulltime	3	2-3	4.00
Female	33	Caucasian	Part-time	5	4+	4.00
Female	44	Caucasian	Part-time	4	1	4.00
Female	46	Caucasian	Fulltime	4	4+	4.00

The site from which I obtained my participants was not my home institution. I chose this institution intentionally so that none of the students knew me. Additionally, they also knew that I had no connection with their home institution or with the faculty or administrators. Therefore, I tried to eliminate, as much as possible, any biasing factor based on a personal association with the institution. I obtained written permission from the college president to conduct my study and informed the vice president of instruction and the deans about the nature of my study. The deans, in turn, notified the faculty, and I received no objections about my study.

Data Collection, Instruments, and Procedure

All my data came from tape recorded interviews. These interviews occurred in a study room located on the second floor of the library at the participants' community college. This location provided a quiet, isolated room that ensured student privacy and eliminated

interruptions. It also made it convenient for students in addition to providing a familiar but neutral environment.

Each interview session began with a brief conversation with the participants to tell them generally that I was interviewing them for a study to complete my doctoral dissertation and to explain that my purpose was not to ask about individual instructors but rather to probe their general ideas related to completing student ratings of teaching. This time also gave us a chance to talk informally about them and to allow them to ask me any questions before we started. I used that time period to have them read and sign the release form and to complete my short demographic survey. Using this time in this manner enabled me to relax the participants to help ensure that they would be candid with me. After obtaining written permission to audiotape their responses, I tape recorded each participant. Interviews lasted from 20 to 50 minutes, depending on how much participants said, with the majority averaging 25 to 30 minutes.

After each interview, I transcribed the tape. Although Glazer (1998) advises to take field notes and not to use tapes and transcriptions of tapes, I decided taping would ensure greater accuracy, would allow me to use the participants' exact words, and would provide a better audit trail if such a verification method is necessary. I followed the protocol found in Appendix A. The participants responded to the following basic questions:

- What do you think about or consider when you are filling out ratings?
- Have you ever not written what you really think about a course or an instructor on a course evaluation, or have you overrated or underrated an instructor in completing an evaluation? If so, would you discuss what caused you to react as you did?

- Under what conditions do you feel you cannot be totally open and honest when completing a rating of an instructor?
- What could instructors and/or administrators do to encourage students to be more candid and sincere in their ratings?

Additionally, with each participant, I asked follow up questions when appropriate to elicit clarifications or elaborations of participant responses to the four main questions.

Once I had completed each transcription, I transferred the transcripts into NVivo and also e-mailed it to the appropriate participant to review. Allowing the participants to review the transcript served as a means for member checking. Two or three of the participants responded by commenting on their speech patterns. Dagmar, for instance, whose native language is German, remarked about the quality of her English. Others commented about how many “uh’s” they have in their speech. However, none of my participants suggested changing anything related to the content. Although Creswell (1998) suggests 20-30 interviews, I discovered that I had reached a saturation point after conducting 12 interviews. All of my participants had indicated that at times they have chosen not to be candid about and/or have overrated instructors.

Data Analysis

I followed the steps of data analysis endemic to grounded theory: open coding (analyzing, studying, comparing, and conceptualizing data into initial categories), axial coding (assembling the open codes in new ways using a coding paradigm), and selective coding (identifying the core category resulting from an integration of the categories from the axial coding) (Strauss & Corbin, 1998). This method of data collection and analysis is an iterative process, one which I continued until I arrived at a saturation point at which no new

information was emerging from the interviews. I used an open coding process in NVivo to code each sentence of each interview.

Once I had completed all the interviews, I began the process of axial coding. I followed the structure that Strauss and Corbin (1998) suggested, identifying the phenomena and then analyzing and sorting the data according to the causal, intervening, and contextual conditions, the actions and interactions, and the consequences. Two common themes emerged relating to what I refer to as climates of fear and trust. The results of this analytical process of coding helped to determine a variety of factors and sub-factors that comprise my study and which appear in chapter 4. Ultimately, through the processes of axial and selective coding, I arrived at my core category concerning the abuse and use of power.

CHAPTER FOUR

RESULTS

In general, the results of this study indicate that a lack of candor and sincerity is pervasive under certain conditions but that other conditions can promote the opposite responses from students. Essentially, student ratings of teaching may do more than provide student opinions about a particular course and the instructor who teaches it when one considers the underlying assumptions based on the climate from which the comments come. Even if the rating instrument is reliable and valid, the responses may not be trustworthy. Results from my study reveal that all the participants indicated that they have either withheld comments or have overrated instructors (or both) and several believe that other students do as well. Conversely, they also cited instances in which they have been forthright or can define conditions that would cause them to be candid. The paradox of possibly receiving untrustworthy responses, either because of lack of candor, lack of sincerity, or both, from students in one course and trustworthy responses from the same students in another course provides the basic focus of this study. The first section of this chapter, therefore, discusses the paradox of candor and sincerity.

From this paradox arise dual phenomena of untrustworthy versus trustworthy ratings. These antithetical responses appeared to be an unresolved dichotomy until I examined the phenomena of these two types of responses by tracing the immediate causes and the ultimate cause of these reactions. This examination, as presented in this chapter, begins with

exploring the factors that cause students to fear or to trust and the instructor's role in creating the climate that leads to these responses. An analysis of the participant responses leads to what appear to be two general classroom environments that cause one of two general climates to emerge. Other environments may well exist, but in my study these two environments seem to account for whether students choose to be candid and/or sincere. One environment creates what I am labeling a "climate of fear"; the other results in a "climate of trust." According to Baker (1992), "Informal day-to-day behavior, with its underlying attitudes and values, makes up that aspect of organizational life that is referred to as the *climate* of the organization" (p. 17). Baker's definition of climate, despite being presented in the context of the organization (the macrocosm), appears just as applicable to the classroom (the microcosm). Which climate occurs in a particular class depends, to a great degree, on the behavior of instructors in relation to their attitudes toward their jobs, toward the evaluation process, and, most importantly, toward the students. For the purposes of this study, I will use the term "climate" to indicate the environment at the micro or classroom level.

Moreover, it seems likely that when one climate predominates among students' experiences, that climate might reflect what could be considered as a "culture of fear" or a "culture of trust" that pervades an entire institution. For my study, then, the term "culture" will indicate the macro or institutional environment. Schein (1985) defines culture as follows:

...a pattern of basic assumptions—invented, discovered, or developed by a given group as it learns to cope with its problems of external adaptation and internal integration—that has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think, and feel in relation to those problems (p. 9).

As Asia pointed out during her interview, students do talk among themselves, and she confirms what most instructors and administrators already know—that instructors carry “reputations” among the students. Additionally, Andrew shared how students in one of his classes discussed their ratings among each other after the class ended. It is such student behavior that illustrates that part of a student culture includes talk that reflects or potentially impacts student ratings or at least preconditions student expectations about a particular course or instructor. Although my results indicate that the instructor assumes primary responsibility for the climate in the classroom, the culture of the institution itself and cultures that exist among students play secondary roles in influencing the climate to a degree and ultimately affecting how students choose to respond on SRTs. Whether student ratings result in candor and sincerity will ultimately depend on whether these ratings occur in a climate of fear or a climate of trust. My presentation of the results, therefore, divides into two major sections comprised of discussions of these two climates.

As I learned from the participants’ comments, fear can manifest itself in various forms. These include a fear of the unknown as well as fears related to the instructor, to the administration, to how instructors may perceive them, and to other students. After a general discussion of fear and the forms in which fear manifests itself, my discussion of the results focuses on the primary force responsible for creating a climate of fear—the instructor. Instructor behaviors can both directly and indirectly affect the climate. They can range from actions related specifically to the ratings and communication skills in general to attitude and personality traits and their treatment of students. Even pedagogical skills and practices appear to play a role, albeit somewhat more indirect, in creating the climate in the classroom. Besides the primary force being the instructor, the administration shares some responsibility

for generating a sense of fear in some students, which I discuss at the end of the first major section. Following this first major section, the second major section presents the second phenomenon of student candor and sincerity in relation to the climate of trust. The presentation of the results parallels the structure of the first major section. Emerging from the findings presented in these two major sections is the core category of power. This chapter closes with a discussion of this central underlying theme: the use and abuse of power.

The Paradox of Candor and Sincerity

Overall, the results indicate that students are sometimes candid and sincere and sometimes not. Therefore, by extension, student ratings may or may not be trustworthy. This apparent paradox begs the question about what could account for this discrepancy.

Most of the participants asserted that they are honest when they complete SRTs. “I’ve always put down what I felt...regardless of what grade I was getting at the time,” said Asia. When asked under what conditions she felt she could not be open or honest on an evaluation, Dagmar responded by saying,

Well...I think, yo...you're asking the wrong person to answer that because I'm sort of very straight forward, and, uh, I'm in a position that I don't care about the consequences, so I don't make myself dependent on grades or on the teacher's, uh, thinking about me. So I'm very honest and there's...I don't think there's anything that would stop me not telling the truth.

Yet even these students, when probed, admitted that conditions could exist that would cause them to withhold their comments. For Asia, it was lack of anonymity. Even Dagmar, who had sounded quite definite about having no fear of commenting, admitted that once she had not commented on an instructor generated evaluation. She described her circumstance this way:

One time I refused with our psychology teacher. At the end of the class, he said, now you have a chance to get 10 points if you write about what did you learn and, uh, sort of hidden evaluation about a horrible teacher. And because I already had my A, I just refused to write the paper and, you know, forgot about the 10 points because I thought it's very biased; it's he collects the stuff and in case he gets, you know, bumped or so, he just shows it to administration, whatsoever, look my students tell me how good I am. And of course they do because they want the 10 points.

As these responses reveal, even those who are likely completely open and honest in most circumstances indicate conditions or have experienced situations that cause them to be less than candid. Despite some variations in particular details, the responses of the participants demonstrate that students will or are willing to be candid and sincere in some circumstances and with some instructors yet withhold their comments and/or their true responses or even overrate instructors in others. To understand this paradox better requires examining both main components of fear and trust.

The Phenomenon of Untrustworthy Ratings

Before I began my research, I knew from personal experience, both as an instructor and as an administrator, that some students write no comments on student ratings of teaching, but the question is why? Are they apathetic or indifferent to the class? Are they afraid to say what they really think? Are they simply too polite to say anything negative? Are they just in a hurry to finish and leave and do not want to take the time to write? Or do other reasons that cause their lack of response exist? Student comments can be useful to instructors to help them understand what is working well and what they can improve when students take the time to respond. If instructors know the reason(s) that students withhold comments, perhaps they can address student concerns and elicit more comments in the future.

The practice of overrating or underrating instructors is more likely to occur when rating instructors using the Likert-scale responses typically found in SRTs. One or two individuals who either like the instructor immensely or, for whatever reason, seek to punish the instructor with low ratings generally stand out among the rest of the ratings as anomalies. These isolated responses likely have little impact on the overall ratings of an instructor or a course. However, if large numbers of students within a give class engage in slanting their ratings in either direction, these slanted results may appear to be trustworthy even though they may not truly reflect the reality of the quality of the instructor or the class. Therefore, knowing how common the practice of overrating and/or underrating is as well as why students engage in it would alert both instructors and administrators about this potential, enabling them to take steps to reduce or eliminate these practices from occurring.

The second and third questions in the protocol aimed to probe the extent to which the practices occur as well as what conditions would cause the participants to avoid being totally open and honest when completing a rating of an instructor. In response to the question about whether the participants have not written what they really think about a course or an instructor or whether they have ever overrated or underrated an instructor, results initially appeared to be mixed, with four participants stating no, two admitting to withholding comments, three admitting to overrating, and three saying they had engaged in both withholding comments and overrating. Of those who remarked specifically about the issue of underrating an instructor, all but one said that they do not engage in that. John said that he has never underrated an instructor, but he stated that he would in a circumstance where he knew that other classmates were planning to overrate a bad instructor. Despite four of the participants claiming that they have not or would not withhold comments or overrate or

underrate an instructor, through the course of the interviews, all four remarked that they had withheld or would withhold comments under certain conditions. Participants therefore fell into the following categories: those who initially claimed they have always been candid and sincere but later revealed that they had not always been candid or that they would not be candid under certain conditions; those who admitted to a lack of candor; those who admitted to overrating an instructor; and those who have both withheld comments and overrated instructors. The results of this study indicate that under the right conditions, all the participants will withhold their comments and/or will overrate some instructors.

Asia, Dagmar, Sara, and Greg comprised the group who initially claimed total candor and sincerity but later acknowledged that they had withheld or would withhold their comments under certain conditions. These participants were fairly quick in asserting that they do not engage in withholding comments or in over- or underrating instructors. For example, when I asked the question, Asia answered with one word: “No.” She went on to elaborate about how she considers writing comments equivalent to when a supervisor does an evaluation of an employee, pointing out the good performance and the areas in need of improvement. “There’s always room for improvement. And usually, I...I write about my teachers very fairly.” But in the very next sentence, she added this caveat: “They don’t see my name.” Later, when asked whether it would make a difference in how she responds if she has any doubts about anonymity, after pausing to reflect she responded as follows:

It might. It might. If I put my name on the paper...and she went and he went through the inter-, in the evaluations, yes...absolutely...because, I don’t know, some teachers are very fair. They just say, you know, this is how this person feels, and that’s that. And some teachers feel that, you know, they take it personal. And so that might change your whole...thing about the course.

Without a guarantee of anonymity, a student like Asia would likely withhold any constructive criticism unless she was confident that she was responding to a “fair” teacher. Asia’s response also illustrates a point that emerges and which I discuss later—that the type of instructor determines whether a student will be candid. Clearly, Asia appears to establish a dichotomy between objective instructors who accept criticism and respond positively and subjective instructors who take criticism personally and, by implication, respond negatively.

In fact, Asia was not alone in modifying her position about total candor with all instructors. Dagmar’s account of the psychology instructor quoted earlier is another example of this phenomenon. In this instance, she withheld comments about how “horrible” the instructor was because she perceived the instructor as having an ulterior motive of obtaining positive comments in exchange for 10 points. Unlike Asia, who would not comment out of fear of retaliation, Dagmar’s withholding her comments appears more motivated by her own personal ethical code but with the implication being that underlying her lack of response was a fear that the instructor might do something to her grade if she were candid. Why jeopardize her A for 10 points that she does not need?

Sara’s answer to the question about conditions in which she felt she could not be open or honest elicited a different response. In her case, she needed to perceive that there was “something in doing the rating” for her personally—but not a grade. Rather, her motives centered on issues related to self-esteem. “I think that...uh...there’d been, that, sure, there’ve been times where I felt that I wasn’t going to be heard... you know, or, um, you know, taken seriously or...uh... you know, honored or respected...” For her, not being taken seriously or respected translates into a feeling that the instructor would not value her comments, therefore

why bother? The fear that appears here is more a fear of rejection than a fear of retaliation. Interestingly, this was not the case for the one instructor she had formally evaluated. For that instructor, she reported that she had “just filled the whole page on the back” with comments. Yet later during the interview, even Sara mentioned a concern about the grade. “And, you know, your, your grade may be...uh... you know...uh... threatened or compromised.” Even if fear of retaliation in the form of a grade is not the primary fear factor, it still appears to be a consideration in the back of students’ minds.

Greg initially indicated that he never intentionally over- or underrates instructors and that he tries to be fair and balanced (“keep it from being, uh, too one way or the other”). Yet when asked about any conditions that would cause him to withhold his comments, he named two: if he thought the instructor would “take it too harshly” and if measures are not taken to assure anonymity.

Fear of being identified appeared in the responses of other participants as well. Of the group that conceded immediately to having withheld their comments on a rating of an instructor, a group that included Margaret, John, Annie, Diane, Susan, and Andrew, fear of identification was the most common reason. What is interesting in some of their remarks is the ways that they believe instructors might identify them even if the instructor does not have access to the individual responses.

Margaret went as far as saying that even though she knew that the instructor would not see the individual forms, she feared the instructor’s being able to identify her based on her phrasing: “...sometimes you can tell just from the wording that someone uses who, who wrote it even if the stuff is transcribed and you can’t read their handwriting.” With the

particular instructor for whom she would write no comments, another contributing factor was that she knew she would have this instructor again because of the program of study in which she is enrolled. John included another way that he felt an instructor might identify him because he withheld certain comments on an instructor rating about a particular incident that had occurred that involved him. A reference to that event would be enough to identify him; therefore, he withheld specific comments about it and modified his comments in general. For Annie, her fear was that the instructor would see and recognize her handwriting and thus identify her. "I think I was afraid that, that they would find out that I wrote them because of my handwriting [laughs]. To think back on it, I think I was a little in fear about it." Like Annie, Diane also feared the instructor's being able to identify her. "If the rating is given by the instructor themselves, I think that would be, um...a case where I wouldn't feel like I could write whatever I want on the paper and have an instructor read it and possibly reflect my grade upon it." With fear of identification running this deeply, it likely does not matter what assurances students receive about the confidentiality of their comments. They do not want to risk being identified. Margaret, John, Annie, and Diane feared negative repercussion from the instructor. The only way that they would feel comfortable writing any comments would be if they could trust how the instructor would take the comments.

For Susan a fear of being identified produced a combination of withholding negative comments along with overrating the instructor.

I have probably overrated... before...if I didn't...feel like...if I thought that professor would be able to piece together who it came from.... Like I'd be afraid to say something negative if I think they're going to see who said it and...where it came from. I'm much more...I much less likely to be critical.

Then when asked if she has withheld comments in that type of situation, she replied that she

left the comments section blank.

Andrew's fear took an entirely different direction. In his case, he withheld comments not so much because of a fear of the instructor but rather a fear of a possible negative impact on his receiving credit for the course. He feared that if the administration learned that an instructor was not doing a good job, it would cancel the course, or if the administration learned that the instructor had missed four of the first eight classes, it would deny students credit on the basis that they had not attended enough of the class. In general, Andrew displayed a "don't-rock-the-boat" mentality. His goal was simply to complete his courses with passing grades so that he can transfer to a university. Therefore, he feared writing anything that might impede his progress. He was also the one student whose attitude was that he would simply endure a bad instructor and then move on ("...once I've gone through it [the course], I'm not worried about it, you know") and that future students should simply endure what he had had to endure ("...hey, you know, if I had to put up this, you're going to have to put up with this, too"). It is almost as if, from Andrew's perspective, having bad instructors is part of a rite of passage for students.

When it comes to overrating instructors, Susan, Jamie, Annie, Diane, Miles, and John identified themselves as engaging in this action. However, fear moves to levels beyond the instructor when it comes to reasons for overrating. Although Susan mentioned overrating a weak instructor because of fear of being identified, those participants who reported overrating instructors did so more because of caring about that instructor than because of fear. Jamie certainly indicates liking the class as being her reason for overrating. "I probably overrated...and that was along with how much I liked the class and the teaching." Annie's response echoes Jamie's. She overrated an instructor who was not organized because she

liked him. “He was very kind, very nice....” However, later in her interview, Annie provided more insight into why she would overrate an instructor:

I think maybe two of the teachers that I had said, ‘Oh, this is, um, something about tenure,’ ensuring that they’d get their tenure, right? So there’s that kind of...um...um...I don’t want to say responsibility...uh...yeah, responsibility to like...and for me I’m always, I’m a really caring person, so it’s like, um, I think that played into, um...you know, my answers being, uh, a little nicer than maybe...truthful. I mean, yeah, I mean, like, um, rating them...um...I don’t know.

For Annie, the combination of being a “caring” person and knowing that an instructor’s tenure was riding on evaluations caused her to feel a “responsibility” for overrating an instructor she cared about. Fear of an instructor’s not receiving tenure underlies the caring that results in overrating.

Although Diane and Miles also claim to tend to overrate instructors, their reasons are different. According to Diane, “I’m not...maybe overrate them sometimes versus underrate them. I’d rather say more good about the person than bad. But, um... I don’t know...maybe I’m an optimist [laughs].” As a result of the entire interview and the impression she gave me, perhaps what she meant to say is that she tries to find the good in people and cares. Her fear may be characterized as a fear of hurting someone’s feelings along with perhaps an underlying fear of something negative happening to the instructor as a result. Caring about an instructor is also a reason that Miles gives for overrating an instructor, but the instructors he overrates are new instructors. Miles was the oldest of the participants and has had teaching experience giving people music lessons. Given his experience, he empathizes with new instructors, realizing that new instructors tend to make mistakes and should not be evaluated as critically because of that. As with Diane, the underlying fear is that something negative could happen to new instructors before they have had a chance to learn to be good

teachers.

Later in her interview, however, Diane also mentioned having to put her name on a rating before and how that impacted her. In response to a question about whether she thought there could be consequences for what she wrote, she responded, “Yeah...yeah, and not even, um...not like the hard set-in-stone consequences to what’s going to happen but just...a different perspective or different treatment as a way she treats you in class the next day after she reads your complaints or something like that.” Diane’s reaction makes it clear that students fear a variety of negative reactions, not just a fear of a lower grade. Instructors can punish students just by treating them differently.

John’s overrating on an instructor also had a basis in liking and caring about the instructor, but in his circumstance, John was concerned about possible underrating by other students.

If I really feel that the instructor was really good and that maybe there’s some people in the class that didn’t like the instructor and I know there’s going to be some negativity there...um...that I might inflate my evaluation a little bit to make sure that the fact that I felt that they were good in the criteria that I was judging on was noticed.

Although he claims to have never underrated a faculty member, John conceded that he would also underrate an instructor to offset the ratings of other students he assumed to be overrating the instructor.

Fear – The Key Factor Underlying Lack of Candor and Sincerity on SRTs

Regardless of whether the participants withheld their comments, modified their comments, or overrated instructors, the emergent theme that underlies all these responses is

fear. The source of the fear varies but falls generally into four categories: what might be labeled as fear of the unknown, instructor related, administration related, and other student related. Consequently, given this number of sources, fear manifests itself in various ways. One can see the variations appear depending on whether the participants were withholding or modifying their comments or overrating an instructor.

Four of the participants specifically used the word “fear” in their responses. Dagmar reports the following:

I know of a lot of students who are afraid of their teachers, and, uh, you know, know that their teachers probably know their handwriting, so in terms of just ticking a box that would be okay but writing a comment a lot wouldn't do it because they're really afraid that would harm them...true or not. But that's the ongoing fear because, uh, you know, the school system is so focused on credits and not on learning that, uh, they never would tell the truth. Why should they?

John specifically uses the word “fear” also in relation to the possibility of the instructor identifying him:

I think...in some circumstances there is a fear that of if you've made something known to the professor and then when it comes time to the evaluation you kind of throw that back out there, that it's kind of...flags you as somebody whose writing something negative because they already know that your feelings on the situation, and when those feelings come out in the review, you're kind of pinpointed as 'Oh, look what he's doing' or 'Look what she's doing.' So...umm...kind of like the fear of being identified even though it's a confidential evaluation.

John used the word “fear” a second time in recommending that ratings occur at the end of the semester after grades were posted,

...so that there's no looming fear of, 'Oh, if I don't say certain things my evalu...er, my grade might get dropped, or maybe if I'm really nice my grade will go up' or whatever. So this way the students don't have any impact on what their grade is going to be once they're given the evaluations, and it's impossible...you know, it's like past the point of changing anybody's mind.

Annie mentioned fear specifically in the context of the instructor somehow being able to identify her. “I don’t want to say paranoid, but I was a little in fear that somehow the teacher, if they somehow saw my comment and saw my handwriting and recognized it that they would be biased against me or...and, uh, somehow grade me down for it.” Significantly, all three participants specifically mention “fear” in association with grades.

Andrew’s fear extended to repercussions for the class as a whole, particularly if the ratings occur early in the semester. As he put it, “And like I was talking about earlier, fear of...you know, maybe something happening within the course. Let’s say everyone complains that, uh...you don’t have enough writing assignments or something, you know and then all of a sudden they packed on more, stuff like that, you know. Or the teacher isn’t organized enough and all of a sudden the teacher goes on like a rampage, you know.”

Even when the participants did not use the word “fear” explicitly, fear is implicit in all the reasons for either withholding comments or for overrating instructors. Fear, however, manifests itself differently among the participants.

Fear of the Unknown.

One general fear emerged from some of the interviews is a fear of the unknown. This type of fear is the most general and abstract because it relates to a general uncertainty that stems from a lack of understanding about the rating process itself. However, this fear appears to provide the foundation for more specific fears related to imagined potential reactions by the instructor and/or by the administration. Furthermore, this fear of the unknown intertwines with the issue of a lack of communication and the role that lack of communication plays in

fostering a culture of fear.

Participants mentioned not understanding the process. For example, Annie said, “You know, I don’t think any of the students really realize, um, what these questionnaires are for.” In her comments about what the instructor or administration can do to encourage student candor and sincerity, she identified two key elements: a guarantee of anonymity and instructors/administrators being “very clear as to what these questionnaires were going to be used for.” The following comments by Sara also reflect an underlying lack of knowledge about the process. “I don’t think you have an idea how seriously it’s going to be taken anyway. You know what I mean? Like, who’s like actually reading these surveys?” Without a clear understanding of the process and the consequences, some students are reluctant to be candid and sincere, or, as discussed later, students develop apathy toward SRTs and would rather not take the time to write comments.

John also mentions students’ not understanding the process. “I think but another problem is, is that...uh...I think a lot of the students may not necessarily be sophisticated in understanding th...the, the real things that happen in an evaluation process.... And I just think that a lot of people don’t understand how they’re necessarily looked [sic].” Without an understanding of the purpose of the process, according to John, some students are likely to use their comments simply to “slam” the instructor rather than make constructive criticisms, and consequently, the instructor and/or administrators will dismiss the comments. As John put it, “But it’s kind of going to discredit a lot of the things that they say in there because it’s such an attack....” Although a lack of understanding of the process could result in students responding in ineffective or inappropriate ways, it can also cause students not to respond.

Finally, Diane also discussed having no sense of the real purpose of the ratings. She described the process this way:

[A]fter we get the evaluations, nothing's ever said...about it again, you know, until the next semester comes and we have another evaluation. But we never hear about, like, the first evaluation or what happens from it or, you know, it's just a piece of paper that we got one day, one semester and...that's, you know...gone and done with.

When students see nothing happening as a result of the process, it can certainly evolve into a general apathy about putting any time or thought into completing SRTs.

Some of the participants commented on their lack of understanding about the rating process itself, the rating instrument, or the results of the process. Jamie did not know how the process works in relation to the faculty. "I don't even know if the person [the instructor] sees it. I don't know if they're allowed to know...what the responses were." Uncertainty about the process extends to the rating instrument itself. Andrew, for instance, suggested that someone should explain what "the impact is going to be on different boxes, you know. How much is it, wh-, how much of a difference is it going to make if I check strongly disagree as opposed to just disagree?" Because Andrew does not understand the impact of his ratings, he reported that, "I usually stick with just the agree and disagrees, not the stronglies. It's like, you know, it's just like I tame it down. I tame it down." Diane also indicated mistrust with the process itself. Although she thought providing instructors with comments is a good idea, she added, "But I don't...really think...um...that they're carried out. I think it's just more filling out the form. I don't really know the process that they go through after the evaluation, but it seems we do the evaluation and nothing ever...happens." Susan expressed similar thoughts about not understanding the process. "I don't know what really happens with them,

so maybe an understanding of the process would [result in more student candor and sincerity]." Later, she added, "But...I, I guess just knowing what the process is would make it more meaningful." Furthermore, in Susan's case, not really understanding the process combined with liking the instructor causes her to worry about the effect a negative rating will have on an instructor. "Well, I think it causes me to worry a little bit with the professors that I like.... You know, 'cause you don't know what's going to happen based on what you say... you know. And I think that...I mean I do think that has an impact [on her choosing to be candid and sincere]...more so with the ones you like than with the ones you don't like...."

Such uncertainty about how the process works appears to underlie student fears of instructor retaliation or fears of nothing happening, which leads to student apathy, factors that I discuss in the next section. When students like Greg see administrators as "just sort of a vague...back there kind of group of people," it indicates that some students have no idea what role administrators play in the rating process. For Andrew, not understanding the process or how results impact instructors caused even deeper fears. "I think I shouldn't...put too bad stuff because I'm afraid maybe...maybe the teacher will get fired and they'll cancel the class. You know, something like that will happen." Later during his interview, Andrew recommended informing students about what the consequences of the ratings will be. "Make sure the students know that, uh...I don't know, it would be nice if you knew that...you know, what's going to happen with this information." In some cases it may matter little to some students what part administrators play; however, as Andrew's comment indicates and as I discuss later in more detail, not knowing the role administrators play can result in a fear of what administrators may do or whether they will do anything at all.

This fear of the unknown appears to stem, at least in part, from a lack of communication. Dagmar used the term “dialoguing.” She contended that from the first days of the semester, instructors and administration should engage in a dialogue with students rather than in a monologue. In her comments about how to improve student candor and sincerity, she asserted, “And I think, uh, one of the key, uh, inputs would be, uh, start communicating. You know, take everyone seriously.” When the system fails to promote a dialogue between faculty/administration and students, students may feel as if no one takes them seriously. It is also a key contributing factor for creating a culture of fear. Dagmar described that fear as something that actually emanates from the state level and filters from the state to the Board of Trustees, from the Board to the administration, from the administration to the faculty, and from the faculty to the students. “So the students is [sic] afraid of what’s above....” As she put it, “It’s sort of a ripple effect, and it has to start somewhere.” If fear exists as pervasively on all levels, as Dagmar imagined, this pervasiveness clearly indicates a culture of fear throughout the institution. Without communication occurring at all levels, it is easy to see how the result could be a fear of the unknown, particularly in terms of how instructors, administrators, and even other students will respond in relation to student ratings and whether students can afford to be candid and sincere. This underlying fear of the unknown in terms of the process itself and how it works translates into more specific manifestations of fear as seen in student fears related to the instructor and to the administration.

Fear Related to the Instructor.

The most frequently mentioned fear concerns a fear of the instructor, that somehow the instructor will know or guess the identities of the students based on the instructor's being present during the ratings and seeing them writing (Jamie), knowing their handwriting (Annie), associating the student with a situation being referred to (John), and even being identified by the phrasing (Margaret). Or participants had a general fear about their anonymity being protected (Dagmar, Diane, Asia). Coupled with the fear of being identified is the fear of retaliation (such as a lower grade) or some other negative instructor reaction (such as increasing student workload) as a consequence for any negative comments (Greg, Susan, Asia, Sara, Dagmar, Andrew, John, Annie, Diane).

Knowing that an instructor may be able to identify individual students impacts whether students feel safe enough to make comments that the instructor may find critical of how they teach. If Greg had any doubts about anonymity, he said, "Then, then I'd hold back." For Susan, the factor that would most influence her is anonymity. One of her first comments related to factors that she considers when completing ratings concerned confidentiality. She expressed it as follows:

The other thing is I'm thinking just how confidential...is it. If the professor's standing right there while you're doing it, they're much more likely to be able to piece together who...is turning in which...evaluation versus if, you know, the professor's out of the room or if you're supposed to submit it in an envelop anonymously later, you know, you're...you're, you're more comfortable, right? So there is like an anonymity concern.

She takes a very pragmatic approach to instructors, saying that everyone has good and bad traits and areas to work on. "But if you say what it is and they know you're the one who said

it, they might...get...they might hold it against you.” With a student like Susan, even though she is likely to point out both the good and the areas instructors need to develop in order to present a balanced critique, she cannot completely trust what the instructor’s reaction might be, and she fears some sort of retaliation. Later in her interview, Susan indicated that she would also generalize her comments to avoid being identified. “I’d probably be leery of mentioning a specific incident that could be linked to me and the comment, right? And I’m not going to say, ‘Well, this happened on this, you know, this day. This is wh-,’ but I will, will try to generalize it so I won’t be identified....” Any amount of instructor presence during the process is enough to create fear of identification for Susan. “...[M]ake sure the instructor isn’t going to be there when it’s going to be collected. I mean, sometimes they walk out of the room, right? But they come back in when it’s being collected. That’s almost the same thing as being there in the room while you’re doing it.”

Andrew described a situation in which student comments led to instructor retaliation:

Like people complained, like, oh, you know...um...you’re not clear enough or something. The teacher kind of gets upset ‘cause the kids are like whining, you know. And the teacher will get weird, pass out this formal, you know, this is the outline of the class and make everything more stringent, you know...a lot more...more difficult.

During his interview, he acknowledged, “I’m kind of worried about the here and now, you know.... so I’m kind of like worried about number one, you know, not really thinking about the bigger picture.” Having experienced what he perceived as instructor retaliation, Andrew’s general reaction now is, “I usually don’t write anything.”

As mentioned earlier, lack of anonymity would likely compromise Asia’s candor.

However, Asia also indicated that fear of being identified is a major consideration for other

students. Early in her interview, she was discussing her own candor because she assumes that the process is anonymous, but her remarks illustrate that even when students like her may trust that the system will protect her identity, still other students are fearful. “You don’t have to put your name, confidentiality thing, I love it because...a lot of students won’t put it because they think, ‘Oh, she knows my handwriting’ or...uh...you know, or ‘She’s going to know it’s me’ or, I mean, ‘She’s going to be upset at us, lower the grade.’” The interviews with other participants support Asia’s statements about other students’ fearing the instructor will recognize their handwriting, identify them, and, subsequently, will lower the grade for receiving negative comments.

Sara discussed the idea of some sort of punishment in relation to student honesty. Her remarks illustrate how the whole issue of student ratings is a part of an entire cultural dynamic that occurs within institutions. Sara begins by describing a particular mentality that exists among some students when their central purpose in taking courses is obtaining a grade rather than education:

I’ve had people tell me, you know, it’s only a semester. If you have an issue, it’s only a semester. You just get out of there, and so, you know, the, the, the, the, the, the adaption [sic] of that philosophy thinking is, you know, don’t...make it a big issue...you know...um...because it’s going to be too stressful, it’s going to be too much hard work, and you may not get anything done. And, you know, your, your grade may be...uh... you know...uh...threatened or compromised. And that’s what you really want, that’s what you, you know [laughs], that’s the goal is the grade. Who cares about the actual in between process of the education? So, you know...um... you know, people will...um...really think like that.

As Sara continued to discuss the mentality of students who are seeking a grade, she indicates that this sort of thinking results from what could be described as an acculturation process that is occurring:

I think there's, you know, all kinds of ways to, you know, uh, get to the grade. It's like, you know, the mouse looking for the cheese, you know? Um, and if, if, if there's a conditioning where they're, like, being punished for speaking the truth or being open and honest, you know, then, then they will go a different route to get the cheese.

Ultimately, the environment can become oppressive. Sara said as much later during her interview.

If, if, if we were somehow...reprimanded or punished or if that became like a problem... you know...um...then...then it's like, 'Well, where do we go?' So it's like, it's kind of like...um...it's, it's oppressive. And so...um...there, there, I think it would have made it even more difficult to be honest on the surveys...yeah.

The implications of what Sara is describing reach far beyond the classroom or even the institution to a general social pressure placed on students to earn good grades. All educators are familiar with those students whose single-minded purpose in school is to achieve straight A's. Several participants mentioned the fact that many students are more grade conscious than learning conscious. Andrew, for example, said, "I don't think the students in the class are really there for like the learning. They're worried about grades and things like that." Dagmar also remarked that, "Instead of really focusing on, you know, education, they [students] focus on grades." Ironically, a focus on grades can also result in students not identifying poor instructors who are easy graders. Dagmar described an instructor who "was horrible. He showed us slides after slides with outdated material from 1988....He didn't even know his own subject....We got our test, take home test where we could look up in the chapters, you know, how the answers were." But she added, "You see, for a lot of students...for a majority students [sic], they more or less are pleased with teachers like that because that means an easy grade. And students don't come really to learn. They come to get a certificate." Therefore, with some instructors, grade-conscious students may fear

retaliation from demanding instructors, but in the case of “easy graders” students may fear a reaction from the administration that could end obtaining their easy A’s. Ironically, while the fear connected with the instructor (retaliation against the student) or with the administration (retaliation against the instructor) differs, the result is the same—students fail to indicate their concerns about an instructor’s inability to optimize student learning.

The causes of such a student mentality of being so grade conscious is beyond the scope of this study to address; however, based on Sara’s observations, the implication is that a combination of factors can result in creating not just a climate of fear but a culture of fear that pervades the classroom. The ingredients include having students who are looking for the easiest route to obtaining an A in a course and having (or having had) instructors punishing them (probably with lower grades) for being candid, combined with the possibility that nothing will happen anyway. The combination of these ingredients can lead to students who enter a course to arrive already having the mindset that if problems occur with the course or the instructor, it is easier just to let it go. As Sara said, “It’s only a semester.” Why bother to endure the stress and to face the possibility that “you may not get anything done”? The result is a possible culture of fear that can manifest itself through student apathy about even trying to affect a change.

If he thought the instructor could identify him, John said that he would modify his comments:

I’d probably be less likely to be as blunt. Umm...if I was to say anything, I would probably...uh...I would probably kind of lighten it up as much as possible so that it didn’t come across...so that the point was kind of still there but no necessarily with the same impact as what I would normally do.

However, if the instructor appears more open to criticism, John said, “I’d probably be more blunt about what was effective and what wasn’t.”

The timing of when the ratings occur in the semester provides yet another factor that contributes to student fear. Some participants were suspicious that ratings coming too early in the semester increases the chances for being identified and/or for some sort of instructor retaliation. A comment concerning ratings coming too early preceded Andrew’s statement above about having a fear of something happening during the course:

But, um, it’s like I just took one for this class, and the class is still going on. We’re not, we haven’t even taken the midterm yet, so it’s like...it’s like the timing, I think. I think it’s better if it’s after the course is over and, um...it’s, I like, that the, if the teacher’s not going to see it, personally.

Jamie also reported an issue with the timing of the ratings:

I do know that I’ve heard other people say that they’re not comfortable filling these out, and it’s too soon to tell. They don’t know what to say. I don’t think it’s too soon to tell. I think you can pretty much answer the questions that they ask within the first three weeks.

Other participants expressed similar concerns about when the ratings occur. In response to the question about the conditions that would cause him not to feel he could be open or honest, John answered, “The biggest thing is the fact that the grades aren’t usually posted...uh...we were just talking about under...about not being able to put what I truly felt...uh...partly when I know that I’m going to be identified. If I know that I’m going to be identified, I may not put what I’m truly feeling.” Asia also commented on the timing of ratings:

This semester I received an evaluation pretty early. Um...it, I guess it’s in the middle of the semester. Uh...I...I think that the...evaluation...I think you should

give time...for students to be in class a little longer, maybe towards...middle end of the semester. I mean, like...the middle of November because you really get to check out your teacher, you know. Um...you don't want to rate them too early because you don't really know...you know.

On one level, she appears to be saying that students “don't really know” the instructor enough to rate that person, but on another level, the implication is that students do not really know how the instructor will react.

Instructors do not always have to present a threat to students in terms of taking retaliatory action in order to generate fear. Sara's concern with being taken seriously and respected reflects a fear of rejection and a fear of being placed in a subordinate role within the classroom rather than as an equal participant. “It's so easy to walk into a classroom setting and, you know, assume a subordinate role because, you know, the instructor has, you know, Ph.D. or, you know, credentials, and you don't.” In a sense, she fears not being in a classroom that fosters critical thinking and an environment that allows students to question their instructors because an environment in which critical thinking and questioning is not happening is the type of environment in which she believes she is not taken seriously. As discussed earlier, not feeling valued would cause Sara to withhold her comments. Later in her interview, she returned to this theme of subordination in relation to being honest in making comments on an SRT. “I'm in a subordinate position as a student [almost exasperated sigh]. You know, uh, no, probably I wouldn't say all the blunt, honest truth...obviously [pause...then laugh]. Because of the process. You know, the process was never given to me, ‘Here, please be honest with me.’”

Instructor-related fears are likely generated among some students before they ever begin a course. As mentioned earlier, instructors' reputations among students can and do precede

them into a course before it ever begins. Andrew mentioned that, “People want to gossip about...the teacher after, after you fill one of those things [SRT] out.” He added that students continued to discuss and complain about the instructor “for a few days. People, whenever you ran into the person in the hall, they would ask you about it.” Under such conditions, it is likely that other students who are not enrolled in that class are overhearing the remarks of the students who are.

Students ask each other’s opinions about instructors. In her remarks, Asia mentioned how she distinguishes between instructors who are “good” because the students will learn but also “work hard” and instructors who are “very lenient” (i.e., easy) for students who want just to “get by” when students ask for her opinion. Asia closed her comments about instructors’ having reputations by saying, “You know, students have rumors.” The word “rumor” is an interesting word choice. For Asia, the distinction is between “demanding/learn a lot” and “easy/not learn as much.” But one can imagine student “rumors” extending to issues relating to instructor behaviors that create expectations about whether the instructor will be retaliatory or will engage in any other type of behavior that would cause students to be fearful. Even the distinction of being a demanding instructor could generate fear among students who either want lenient instructors or who lack self-confidence about their abilities to succeed. The “rumor mill” can also reach the ears of the instructors, which may impact how they treat students.

Both Asia and Margaret referred to instructor reactions based on student comments. Asia commented on indicators that instructors know their reputations based on comments they make to students in class as well as talk among students:

They might say little comments like the first of class, you know. They might say that, you know, or bring up a story from a previous student.... They sometimes think that they are too hard or they grade too hard...or from the previous rumors going around or whatever. Or they're unfair, so they change it up. So I think they go about it in their own little way for rumors 'cause students...different teachers have...uh...reputations. You hear from fellow students. So I'm pretty sure...she's hearing it or he's hearing it.

Margaret cited the example of one of her instructors whose reaction to student comments created a fear of writing comments. As Margaret explained it, "In the particular case, the instructor had sarcastically quoted, um, comments from previous evaluations to us in class...at...repeatedly. And so I, I censored myself because of that." One can imagine that these "rumors" and instructor reputations can easily spread a climate of fear, which is only enhanced when instructors like Margaret's refer to student comments or bring up their own reputations in class. The participants did not restrict their fears to the classroom or the instructor. Fear stretches to the level of the administration as well.

Fear Related to the Administration.

Manifestations of students' fears in relation to the administration emerged, paradoxically, in basically two diametrically opposite forms: fear of administrative action (that the administration will hurt the students or the instructor) or, conversely, fear of administrative inaction (that nothing will change). In the first case, those participants who feared some sort of administrative action withheld comments or overrated instructors out of a sense of protection whereas those who viewed the administration as inactive withheld comments while displaying a sense of apathy.

Andrew's reaction provides an example of how student fears can reach a state of paranoia and illustrates the lack of understanding that some students have about the purpose

and ramifications of student ratings:

When the teacher had an attendance problem, I didn't say anything because I thought maybe if you write down, oh, the teacher hasn't been here for four out of the eight times we've met, then the district gets and says like, well, this doesn't count as a class anymore. You know, it's like you weren't in class enough or something, and then maybe they would drop the class.

It is one level of fear to expect the instructor to make the class more demanding as the result of student comments; it is another to assume that the college will cancel a class and/or deny a student credit.

While student silence appears to stem from a fear of some sort of retaliation occurring, the fear underlying the overrating of instructors presents a different set of fears. Although both Dagmar and Susan cited fear of instructor retaliation as being a factor in overrating, more common is the fear that something negative will happen to the instructor based on student comments. The fear, then, concerns a perceived potential for administrative punishment of the instructor or even of the students in the case of Andrew who feared a course cancellation.

Annie describes a situation in which she overrated an instructor that she liked. She described this instructor as being "very kind, very nice"; however, his lectures were disorganized. "There was a question, how is this teacher's...or, or is this teacher's lectures organized. And I, and I think I put, like, very...very well organized and they weren't really [laughs] that organized. Umm...so, yeah, I was definitely bias [sic]." Although being kind in return to an instructor she likes may have been Annie's primary motive for overrating the instructor, later in her interview she makes it clear that such an overrating may be necessary to protect that instructor from negative consequences:

I think would tend to go higher because I would wa...I, you know, in my mind I thought, well, I wasn't really exactly sure if these questionnaires, you know, how important they were and, uh, what kind of impact they'd have on the teachers, but, um, in a way I, I think I thought, well, you know, I want to rate them high because I...maybe that'll ensure that they'll stay here, you know. Their jobs will be secure [laughs].

Annie was not alone in wanting to avoid hurting an instructor. Susan echoed her sentiments but with a slight variation in the conditions under which the overrating occurs.

Susan explained, "I've recently had this situation, and I...where you like the instructor...but there's a couple of things they don't do well. And so you don't want them to get...in trouble...for not doing their certain things well.... Um, when it's just quantitative, you feel a little guilty, I think, making a negative comment because you can't explain why it's negative and you don't know how that's going to affect them as a person." For Susan, however, she implied that she would be more honest in completing the quantitative ratings if she has room to explain her quantitative ratings to emphasize that "it's not that they're bad as an instructor." Even when students do write comments, they may still modify them out of a fear that something bad will happen to the instructor as the following comment by Sara illustrates: "But I think the...the, the idea that I had in my mind was that, you know, um, you want to be as diplomatic as possible because it may affect it [an instructor's receiving tenure] negatively."

Also admitting to overrating an instructor because she liked that person, Jamie said, "I mean I've had some unorganized teachers. I know there's a question about organization, and I...I probably said whatever the best one is, excellent or whatever it is even though maybe they were not so organized just because I like them or something." Despite not specifically mentioning that she overrates the instructor to protect the instructor, the implication seems to

point to this as being the motive rather than the motive being to reward an instructor that she likes. She did not say, for example, that she automatically gives an instructor she likes all high ratings.

As these three examples exhibit, the combination of liking an instructor and of fearing that the administration will punish her or him in some way along with feeling guilty about giving a negative rating can result in overrating. Additionally, although Diane did not mention having guilt feelings when she commented about tending to overrate instructors because she is “an optimist” and would rather say more good than bad about an instructor, I wonder, in retrospect, whether she prefers saying more good than bad about instructors to avoid feeling guilty.

Another occasion could elicit overrating from students to prevent instructors from losing their jobs—when instructors announce to students that the administration reviews the ratings for tenure purposes. Andrew described how knowing this could impact his responses:

You know, I don't think that's ever happened, but that could happen, like, oh, my job's on the line...er not, maybe not that extreme but...they're like, you know, reevaluating my performance, you know, something like that. And it would, it could affect me...you know. So I think that they should be a teacher still, I'll give them good marks, you know...even if they don't deserve them so like, you know, not being honest on the sheets.

In this type of situation, an instructor could make comments to elicit student sympathy.

Andrew's remarks illustrate that he is the type of student who would overrate an instructor under these circumstances.

Miles, the oldest of the participants and a musician who has given people music lessons, provided the one alternative reason for overrating an instructor:

I guess if the instructor's new because I, you know, like...I, I've been a teacher, and I've been in that position. And I understand the pressures that it might involve, and so if a, if an instructor's new, I...I kind of feel like sometimes I want to give him or her a break because, um, they're...they're breaking...they're...it's their maiden voyage so to speak.

Clearly, in Miles's case, he empathizes with new instructors, having been one himself at one time. Still, the implication of Miles's response for new instructors implies that possibly he is overrating to ensure that nothing negative happens to the instructor as a result of lower ratings.

The inverse of fear of administrative action is the perception that some participants had that administrators are likely to take no action. Students can perceive the administrative role as so removed from the students and perhaps even the process that they have no impact on what an instructor does. According to Annie, "I know administrators are kind of, uh, from my experience seem very removed from the students...um...in the sense that they have their own, kind of, separate work. You know they're working in the school. They're...they're not really involved in student services or, um, they're at a different level than the teachers." Greg presented another example of students' not really understanding the role administration plays. When I asked whether he had any expectations of administrators in terms of what they should do, he replied as follows:

Yeah...if the teacher is very good or very bad, then the administrator should...act accordingly. But...um...I, I don't...know how, how much administrators administrate... [laughs] or exactly what they do in how their role in the, uh, the teachers'...uh...job is.... Yeah, I mean, they're back there, and they, they control...I think they control policy and...uh...but they aren't, they aren't, you know...teaching, and they don't have the direct relationship with the students. So I, I don't...know...they're, they're just sort of a vague...back there kind of group of people.

The sentiments that Diane expressed likely speak for a certain group of students who doubt that anyone will do anything. She describes her view of the rating process this way:

Um...I think that it's a good idea to get the feedback as to how the faculty is doing and how the teachers are doing with their students and such things. But I don't...really think...um...that they're carried out. I think it's just more filling out the form. I don't really know the process that they go through after the evaluation, but it seems we do the evaluation and nothing ever...happens. We just do an evaluation.... It's just kind of like, you know, you don't take them as seriously. You don't...you're not...you're only going to do something a certain amount of times and have nothing come from it and then do it again with the same full...fullness that you did it the first time, you know. So...I think a lot of people just don't take them that seriously. They think, you know, whatever. Like what I strongly agree versus I strongly disagree, it's not going to matter in the long run.

Although Diane does not specifically name the administration, the fact that Diane sees nothing changing implies that lack of responsiveness to student comments extends beyond the instructor to the administration as well. Clearly, Diane's remarks reflect apathy setting in when students fail to take the ratings seriously and develop an attitude that "it's not going to matter in the long run."

Similarly, Susan mirrored Diane's sentiments. "I mean, I can't say sometimes in your head you might be going, 'Well, is anyone going to be paying attention to this,' right? And so I can see, and that's, you know, you might just...not put much thought into it. I don't; I tend to be pretty conscientious, but I could see other people...doing something lightly...just kind of filling out average or whatever because it, it does sometime you kind of do wonder, does anything ever happen with these...anyway." Finally, as discussed earlier, Sara's comment about students asking her "Why bother?" because "you may not get anything done" illustrates a sense of apathy that results from the perception that no one, neither the instructor nor the administration, will do anything to address the perceived shortcomings.

This uncertainty about administrative involvement and the fear that no one will do anything anyway likely contributes to student apathy about completing SRTs. The preceding comments from the participants support this conclusion.

Fear Related to Being Perceived as Biased.

Two of the participants expressed another reason for withholding their comments that might best be classified as a fear of being perceived as being biased. This type of fear is a variation of fears related to the instructor because it is the instructor's reaction that these students want to avoid. However, that reaction may or may not result in an impact on the student's grade. The reaction may simply cause the students to lose respect in the eyes of the instructor.

Andrew described a situation in which he felt the instructor concentrated too much on ethnic and gender issues, yet he would not express his views on this in the comments section on the rating. "I won't say anything 'cause it's like, you know...maybe I'm just trying to be, uh, like racist or whatever. You know, I'll be afraid to say something like...I sometimes I wish we'd study more like classic stuff instead of this, you know, kind of new, different ideas and things."

Susan mentioned an experience in which she had instructors who were non-native speakers of English. She found them difficult to understand even though she said they may well have been good instructors. "And maybe in some of those cases I might even...over value, rate them because of that because I don't want to be...prejudiced against them just because they're from somewhere else."

In these examples, the fear of being perceived as being biased or prejudiced caused

these students to withhold comments or overrate instructors.

Fear Related to the Students.

Fear can evidence itself even at the level of fearing how other students will respond on their evaluations. Although this was not a common theme among the participants, this seems to be the implicit message with John who would overrate an instructor to counterbalance potentially negative comments from other students. “Like personally, if I...well if I like an instructor and like I think I would be more prone to overrate them if I knew that other people were challenging the instructor but I believed in the standards that the instructor was enforcing.” In this instance, the fear is that the administration will somehow punish the instructor as a result of student ratings, but unlike the situation in which the participant is trying to protect the instructor from the participant’s true ratings, here the participant is trying to protect the instructor from the ratings of other students.

Instructors and students like John each suspect and fear that some students may try to punish instructors through low ratings in retaliation for a low grade. Andrew also commented on this occurring in the context of some students being in class for the grade rather than for learning:

Yeah, it’s kind of like pess-, they’re like kind of negative, like the teacher’s a bad guy, you know...not the teacher’s there to help you. It’s more like, oh, I need to make sure I get a good grade. The teacher’s going to give me a bad grade...you know. They’ll just kind of...they’ll rip them on the questionnaires.... [I]f they think that the class is difficult...if they think it’s too hard, they’ll give them poor marks.

When asked whether he had ever done this, he answered no, but he said that he knows it happens because students discussed their ratings after class among each other. Therefore, if

students discuss their intentions concerning their ratings before the ratings occur, it is conceivable that a student like John would overrate an instructor to counterbalance the intentional underrating from other students.

Annie also commented on students who are not serious about learning. In the case of these students, she reported that she “can only assume how they vote.” Then she added, “But you know that the victim role of blaming the teacher for not succeeding and doing well in school, so I’m sure that there’s a, there’s a substantial amount of people that fall into that category.” When asked if knowing this about other students would influence her to overrate an instructor, she said that she thought she might.

How common this practice is, based on this study, is unclear. Only two students even mentioned the possibility, and only John admitted that he might underrate an instructor as a counterbalance to students who may be overrating an instructor. John described the circumstances this way:

[I]f I didn’t feel they were as good and they lacked off in some areas but I know that some students just adored the professor because...whatever reason they were getting...doing very well in the class when I felt that...things weren’t optimal learning environment, but I might lower my ratings a little bit to kind of compensate for how I felt other people might be judging.

Although Miles said that he had “underrated” a biology instructor (discussed more fully in the *Communication Skills* section below), in that situation a low rating appeared justified given the instructor’s treatment of students. Given that context, it appeared that Miles treated the term “underrated” as being synonymous with “gave a low rating” rather than as meaning “gave a lower than deserved rating.”

The possibility exists that some of the other participants underrate instructors for this or

other reasons, but that this practice may not be something they wanted to admit. However, Margaret, Greg, and Jamie specifically said that they do not underrate instructors. Presenting an example of an instructor she did not like “personally” and another instructor whose views she did not agree with, Susan asserted that she gave both positive evaluations as instructors. As she put it, “Um, you might not like them...as a person because you disagree with their stan-, viewpoints politically or religiously or whatever it is, but that doesn’t make them necessarily a bad instructor. It’s just that you wouldn’t necessarily...want to be friends with them.” Andrew did not specifically say that he did not underrate instructors, but he implied that he did not when he said that he is always honest on the Likert scale factors and that he would simply withhold his comments rather than expressing anything negative. Diane also viewed herself as being “pretty honest with the evaluations.” Finally, Asia bluntly answered, “No,” to the question about whether she had ever withheld comments or over- or underrated an instructor, and she later added, “I feel like I can, I can be myself and tell the truth,” (as long as the ratings are anonymous). Therefore, overall, the participants as a whole gave no indication of trying to conceal ever intentionally underrating an instructor.

As is clear from the preceding results, fear can arise from a variety of sources including a fear of the unknown concerning the process itself and of what action or lack of action administrators will take. Among these factors, the most important one for creating a climate of fear in the classroom is the behavior of the instructor.

Instructor Behavior and a Climate of Fear

Through their interviews the participants referred to a number of instructor behaviors, some specifically related to student ratings and others related to their relationship to students

as well as related to their teaching. What emerges from these comments is a sketch of some of the factors that likely contribute to creating a climate of fear in which students avoid being completely candid or sincere when completing SRTs. These behaviors fall into five general categories: actions related specifically to the ratings, communication skills, attitude/personality traits, treatment of students, and pedagogical skills and practices.

Actions in Relation to the Ratings.

One sure way of ensuring establishing a climate of fear is for an instructor to make sarcastic remarks about prior student ratings or comments to the class. Margaret experienced this circumstance, as cited earlier, when an instructor quoted comments from prior evaluations sarcastically. In turn, Margaret said that she “censored herself.” What compounds the problem in Margaret’s case is that she is in a specialized program in which students take more than one course from the same instructor. As a result of the instructor’s comments about prior ratings, Margaret said,

The teacher that I talked about that I felt that there would be retaliation from, I really like, and I don’t think that she’s a great teacher. I think she could really improve, but I’ll never write. I’ll have her again. She’ll have an effect on my career forever, and I’ll never write a comment on an evaluation for her.

Although Margaret liked the instructor personally, the instructor generated enough fear of retaliation that Margaret chose never to write comments about this instructor even though she acknowledged elsewhere during her interview that she is the type of student from whom “...lead just pours out of my pencil” when an instructor is receptive to student comments.

Margaret’s experience with an instructor making comments about student ratings was not an isolated incident among participants. As detailed earlier, Asia commented about

instructors making comments in class. “They might say little comments like the first of class, you know. They might say that, you know, or bring up a story from a previous student.”

Such comments likely undermine students’ confidence in the system’s ability to have a positive impact on creating any changes in the instructor, and they help create a climate of fear, perhaps causing students to wonder whether instructors see the original rating forms that contain the students’ handwritten remarks.

Several participants also commented on the effect that the instructor’s remaining in the room has or would have on their willingness to be candid or sincere. Diane said,

If the instructor wasn’t present in the room, that’s definitely a more comforting scenario without writing about somebody who’s not there.... If they’d stay in the room or if, um...I think that would be.... I think it’s...it’s more like the, they know what you’re about to write down, and it’s...you know, it...a teacher has a certain amount of auth-, authority over her students, you know, as to what they’re going to do with the rest of their lives and what grades they’re going to have, you know, and to go on to whatever level that they want. So I think it’s a little bit...intimidating...to have the person that you’re writing about to be in the room and be watching what you’re writing or, you know, seeing the form that you’re filling out, especially if your names are on there. I believe last time we actually did put our names on the evaluation form.

Diane was not alone in commenting on the instructor’s presence during ratings generating fear. Susan, in a comment about overrating quoted earlier, mentioned her fear of saying anything negative “if I think they’re going to see who said it and...where it came from.”

Later during her interview, she added, “And if they’re not the ones doing the collecting and they’re not the ones there, then I’m much more comfortable being honest.” Another factor that emerged from Asia’s interview also accounts for student fear of being identified if the instructor is in the room even though she was commenting on the fact that usually another instructor conducts the student ratings. “So you don’t have to worry about, well, um, who’s

writing more...like who's the last person to give their evaluation back and that they wrote more. She's going to know or he's going to know." The implication here is that the course instructor's presence during ratings would create worry about being identified simply based on the length of time a student writes in connection with the length of the comments.

However, as I shall present later, an instructor's presence does not automatically mean that students will withhold their comments. The key element is whether students fear or trust the instructor. In the examples cited above, the students had reasons to fear being identified. A climate of fear does not suddenly emerge at the time students complete their ratings of the instructor and the class. When instructors make sarcastic remarks in class about prior ratings, that behavior illustrates one way of creating a climate of fear that directly relates to the ratings themselves. Fear of an instructor's presence during the ratings signals that a climate of fear already exists prior to the ratings. Other factors, therefore, also contribute to creating a climate of fear.

Communication Skills.

One of these other factors revolves around how instructors communicate with students during class. As discussed earlier in relation to Dagmar's comments, lack of communication is a key component for generating a fear of the unknown. A classroom environment based on a "monologue," to use Dagmar's word, discourages students. "It's not a monologue, or it's not a...you know, the teacher or administrations are the dominating ones, and the students are the ones who are stupid and who need to be filled up with, you know, nonsense knowledge. That's part of education, but, you know, not all. Education's really happening in a dialogue." Sara's remarks, as discussed earlier, point to a lack of dialogue contributing to her not

wanting to write comments. When instructors fail to establish dialogue with their students, the consequence is likely to be the creation of apathy and a climate of fear. However, instructors can generate fear in other ways involving communication.

Besides any sarcastic remarks that instructors may make in regard to prior student ratings, instructors can create fear by ignoring student questions in class or cutting off students' comments when they are saying something with which the instructor disagrees. Miles described an instructor who ignores students. "There was a, one instructor I had that was a biology instructor, and...um...he was basically a good instructor, but he didn't like to answer questions, and I, I would try to get my questions in. I, even if I raised my hand and said, 'Excuse me.'" Not answering questions was not this instructor's only problem with communication as is clear in the following account of this instructor's behavior in class:

This particular biology teacher would talk, and he would kind of turn around so you couldn't really hear. You'd lose his voice every now and then. And when some of the students in the class, you know, uh, complained, he would just say, 'Hey, look, you know, um, I'm here, and, you know, if you, if you can't hear me just, you know, you need to follow me around,' you know, things that...condescending remarks that, you know, just didn't add to the, the learning environment itself.

The result, according to Miles, was that by the third week of the semester, 15 students had withdrawn from the course. Miles said that he underrated this instructor.

Annie presented the example of an instructor who cut students off. "I respected her, uh, very much, but, uh, she, um, would cut people off during discussions, you know, like we would talk. She would, she would ask questions. If she didn't like your answer, she would kind of cut people off." The result for Annie with this instructor was that she said she would not write comments for fear of being identified. This instructor's actions created a climate of

fear, at least for Annie.

When instructors are not native speakers of English, instructors have problems communicating with students on a different level. If students, such as Susan, are afraid to say anything for fear of seeming prejudiced, the fear exists even if the instructor may be trying his or her best to teach well. The real problem is that if students truly cannot understand non-native speakers teaching a course, according to Susan, “maybe that really *does* make them a bad instructor in the, you know, in the context.” Yet if students are afraid to write comments to that effect out of fear of appearing prejudiced, nothing will change.

Being able to communicate with students extends beyond the classroom as well. In discussing factors that she considers when completing ratings, Jamie mentioned the ability to talk with the instructor outside of class. “A lot of times I don’t speak...mmm...I speak in class plenty, but if I need to speak to the teacher after or in private, I like to make sure I feel comfortable with the teacher even if they’re inviting to that.” Besides how communication skills influence the classroom climate, the attitude and personality traits of the instructor also contribute to creating a climate of fear.

Attitude and Personality Traits.

Instructors can convey their attitude toward teaching and toward students in a variety of ways. These are factors that influence students’ views of instructors and classes. Jamie actually uses the term “attitude” in describing two general types of instructors:

I guess their attitude towards teaching, and, um, you know like so far all of my teachers are really...if I turn in something and there’s something wrong, they’re reall...good about explaining what’s wrong so I can change it next time whereas I’ve had teachers that just...it’s wrong and don’t...express what to do to fix that, which really makes me not, I mean, I don’t really consider that...teaching...really.

An instructor who merely marks the errors but fails to help students understand and correct those errors likely conveys an attitude of not caring about students.

A characteristic that combines communication skills and attitude is approachability. Susan commented that approachability is “really important” for her, and Jamie implied the same characteristic when she related her need to “feel comfortable with a teacher.” Additionally, ways that instructors communicate also convey their attitude toward students. As already discussed, making sarcastic and condescending remarks to students as well as ignoring students’ questions both convey a basic disdain that easily can result in the creation of a climate of fear. Additionally, participants provided other examples of negative instructor attitudes.

Participants sometimes distinguished between instructors who care for students versus those who do not. As Miles described it, “I feel that...the teacher should be there for the students. I mean, that’s the main reason, uh, even though there’re a lot of teachers that I feel...that I’ve experienced, uh, that are just, just there for a paycheck.” An instructor can convey this sense of teaching’s just being a job if they lack enthusiasm. Dagmar describes how new instructors often come in being enthusiastic and excited about teaching, but eventually they lose that enthusiasm. Although she qualified her statements to say that not all long-time instructors are this way, “there are those old teachers, tenured for years and years and years, and they [don’t] give a damn about students. You know, [sigh of frustration], and you can’t do a thing about it. They don’t keep themselves [sic] up with the, you know, the psychology teacher I had. He was horrible.”

Greg admitted how sometimes he just does not care about the instructor or the class. As

a result, he said, “I think they’re doing an adequate job, and I just want to circle the numbers and go home. An, and it’s really, I don’t...have any real input or anything.” He went on to say that it was a lack of interest in the class that caused his reaction. Greg did not say that he feared writing comments; however, his lack of interest could reflect the inability of the instructor to generate any interest in the subject matter in him.

Another characteristic that promotes a climate of fear involves having no compassion or understanding of students’ needs. John described this inflexibility as being a negative characteristic of some instructors when discussing what he notices about instructors. “Are they a task manager that’s just, you know, this is the rule? You come in, this is the time you take the test, this is the homework, and there’s [sic] no exceptions to anything. This is the way it is, and they’re not flexible to the needs of the students.” Interestingly, John is just as critical of instructors who have no rules. He continued the previous statements adding,

Or are they, you know, maybe some teachers are way over-catering to students in some ways and bending over backwards to do things for people and getting taken advantage of? And I think both of those cases are negative because it doesn’t teach the student to take responsibility for themselves, but I think we also need to encourage instructors to really take into consideration that there are going to be exceptions to the rules every once in a while.

As referenced earlier, the biology teacher that Miles talked about caused 15 students to withdraw from the course within the first three weeks. “I believe it was due to his attitude and how he approached teaching itself.” Based on comments like this and those cited earlier, clearly the way instructors approach teaching and how they deal with their students conveys their attitude. Presenting an overall negative attitude to students appears to contribute to creating a climate of fear in the classroom. It is also apparent in how instructors treat students.

Treatment of Students.

How instructors treat students does impact how they respond on student ratings. For instance, how instructors treat students is one of the first factors that Jamie said she thinks about when completing a rating. “I watch the equa...like, I don’t know if it’s equality, but how the teacher...treats everyone in the class. I know if it maybe it has to do with I’m a little bit older, but [laughs] I always pay attention to that.”

I have already quoted Dagmar concerning how instructors tend to engage in “monologues” rather than in “dialogues” with the students. As she described it, “...the students are the ones who are stupid and who need to be filled up with, you know, nonsense knowledge.” When students feel excluded because instructors simply lecture to them to “fill them up with nonsense knowledge,” students like Dagmar feel as if these instructors are treating them as being “stupid.” Sara’s comments, quoted earlier, also point to student treatment.

Sara expressed feeling as if she was not going to be heard, which led to feelings that the instructor did not take her seriously and did not honor or respect her. Such treatment has a psychological impact on students, which Sara herself makes mention of. “...[S]o, you know, um, you know, it, the...the self-esteem issues that, you know, the psychological issues are, you know, sort of play in...into, you know...how is, what’s my incentive for doing this?” Like Dagmar, Sara mentioned feeling as if some instructors treat students as subordinates. (Such behavior can also occur outside of class. An instructor related the following incident to me. The son of this instructor saw one of the instructor’s colleagues in a coffee shop, and approached this instructor by saying, “You’re Jane Doe, aren’t you?” Not recognizing the

student as the son of a colleague, the instructor replied, “It’s Dr. Doe.” Although this was not an incident that any of the participants related, it illustrates the kind of behavior that the participants can experience to make them feel “subordinate” or inferior to an instructor.) In Sara’s case, an instructor’s treatment of her can affect her self-esteem, can affect her psychologically, and such treatment can lead students to question what their incentive is for making comments or being sincere.

The instructor that Annie described who cut off the students when she did not like their answers or Miles’s example of the biology instructor who made condescending remarks to students who complained that they could not hear him provide other instances of how instructors can belittle students in class. The subtext of the message might be, “You are not important. Your opinions do not matter.” Such a message translates into fostering a climate of fear in the classroom.

Pedagogical Skills and Practices.

Despite not mentioning pedagogical skills and practices directly in relation to creating a climate of fear or trust in the classroom, participants did mention them in relation to factors that they consider in rating instructors. One practice involves the organization of the instructor combined with the pace of the course. Asia is a student in a vocational program that covers extensive amounts of information, some in eight-week classes. For her, the issue is not the amount of material but how well the instructor has prepared her. “So I find that if the teacher is...is not organized, is not presenting the information clearly...uh, that I get frustrated. You know, I feel like, you know, she hasn’t or he hasn’t prepared me well, and now he or she is expecting me to know this. And I feel that that’s unfair.”

Students can easily evaluate how organized the instructor is by examining how well the instructor follows the course syllabus. Sara mentioned this factor in her comments as she brought up how some instructors may have a lot of knowledge but also introduce “irrelevant information and random knowledge.” She even pointed out “...like the syllabus, you know, like that’s supposed to be it, the contract, you know, the focus.” Following a syllabus is so important for Sara that she also added, “I don’t know if there’s, you know, some sort of, uh, class [laughs] to teach how to write a syllabus, but if there isn’t, there ought to be because that really would be, you know, the, the, the, um, the place to start.” Sara followed these remarks with an example of an instructor who came to class and admitted to the students that she did not have a syllabus because it was her first time teaching this particular course. At the end of the story, Sara concluded with a question. “You know, are you [referring to the instructor] saying that you just don’t care, or...[laughs and does not complete the statement]?” This question demonstrates how students can view an instructor’s pedagogical practice as a reflection of the instructor’s attitude toward teaching or toward the class.

Instructors can also convey an attitude of not caring by engaging in the practice of failing to return assignments in a timely manner. Diane brought this point up, saying that it is not the amount of work an instructor assigns that is the factor. She related the situation this way:

Sometimes I think teachers will assign more work than, uh...they’re willing to correct and do, and our assignments won’t...like if I do a test I expect it...my test would be handed back within that week. And so I think if a teacher doesn’t hand back a test or an assignment with the week that you turn it in, I think it shows that she’s...not as...let me see how to put this...um...not as dedicated to correcting the work as she is to assigning it. And I think that’s a big factor.... I don’t mind, uh, the amount of work as long as I feel I’m getting feedback from the work getting, returning it.

Related to the instructor who does not return work in a timely manner is Dagmar's example, cited earlier, of the instructor who continues to use outdated slides and old lecture notes year after year. An instructor who does not keep his or her course materials current also demonstrates a lack of caring and impacts how students perceive the instructor's pedagogical practices.

Administration, the System, and a Culture of Fear

The participants made relatively few comments about the administration and its role in the process. As discussed earlier, participants' remarks about administrators reveal a broad range of views about the part they play. Some had no real idea about what the role of administrators is while others see administrators doing nothing. Miles even believes that how an instructor teaches is not the job of the administrator. "I don't, I really don't think it's...uh...up to the administrators to do any of these things. Uh, I mean, you know, to try to make the teacher work." On the other end of the spectrum are those who fear administrative reaction.

Some like Dagmar and Sara consider administrators an integral part of creating an institutional environment that promotes communication and dialogue between students and faculty/administration. Dagmar described what for her would be the ideal environment that would begin in the first days of class:

Administration, faculty could manage to give that sort of attitude or make that attitude transparent with students, that they're really together in the same halls, in the same building where knowledge is, uh, exchanged, where the students are heard as well with his/her own expertise and knowledge and wisdom, then, you know, every student would be glad filling out an honest evaluation.... But as long as the students is [sic] in a position, umm...in, uh, an oppressed position, it never

will happen.

Sara's comments also indicated that with more pervasive communication would extend honesty throughout the college, not just in the classroom:

That process of honesty, you know, where it applies to surveys or, you know, writing or architecture or education or whatever, um, you need, we need to teach students to be honest from the very beginning...you know, and administrators to be honest from the very beginning and staff to be honest from the very beginning and employees to be honest.

As Dagmar and Sara's comments above show, participants spoke of conditions that have promoted or would promote candor and sincerity in completing SRTs. Just as the results of SRTs may be untrustworthy, so, too, can they be trustworthy. Just as factors exist that contribute to creating a climate of fear, other factors exist that contribute to creating a climate of trust. These factors, again, trace back to instructor behaviors and attitudes.

The Phenomenon of Trustworthy Ratings

Despite the admissions from the participants that they would withhold their comments and/or overrate or underrate instructors or had done so already, overall, they also indicated that they try to be objective, fair, candid, and sincere on most occasions when they rate instructors. The fact that the participants' ratings can be either trustworthy or untrustworthy depending on the instructor and the circumstances influencing them at the time they complete the rating indicates that trustworthiness is relative rather than absolute. The participants made it clear that they have no compunctions about providing objective comments under the right conditions. All the participants gave me the impression that they would rather be honest than to withhold their comments, and without a sense of fear, those who overrate instructors

would also likely provide more honest ratings. The participants' comments support these conclusions.

In relating what she considers when completing ratings, Dagmar called the process "my responsibility to give my input in, you know, developing education." She viewed it as her chance to provide the instructors with comments about how they are teaching, what they can improve, and how clearly they are presenting the subject matter. Her description of what she considers included an account of how she uses the process to evaluate herself and what she has learned as well. For her, ratings are a responsibility she takes seriously. So when she said that "I always write what I think, uh, to the best of my knowledge and to the best how I would evaluate myself because without evaluating myself I really can't evaluate the teacher's teaching style or the subject or whatsoever," it is clear that she views the process of rating a course and an instructor as an important duty. Furthermore, she measures an instructor's effectiveness in relation to her self-evaluation of her own learning, a characteristic that would likely lead to more useful comments for the instructor being rated. Under most circumstances, I would expect Dagmar's ratings and comments to be completely trustworthy. Other participants gave me the impression that they are just as sincere about their trustworthiness when they have nothing to fear.

Similar to Dagmar, Margaret uses her comments on ratings as a way to reflect on her feelings about the course and the instructor. Although she adamantly refused to write comments about the instructor from whom she feared retaliation, she said that usually she has no problem with writing comments even if she thinks an instructor will not pay attention to them. "I'm inclined to still give them the comments for two reasons. One, in case they do

[consider them]. Three. One in case someone who hired them is reading comments, and then the third is, um, it's a way for me to, to think about my feelings and maybe get some closure on that." Without a fear of retaliation from an instructor, Margaret takes writing comments quite seriously as her following comments demonstrated: "I've certainly had instructors that I have felt would just, wouldn't really care what comments I wrote, and I wrote them anyway. And I try to write, you know, I try to think about what will help the instructor in the future, not just I don't like when you wear pink dresses or something. But, um, or I don't like you or this class sucks. That's not helpful."

Trying to be honest in a helpful way characterizes the statements of other participants as well. Take, for example, this remark of Sara's: "It's kind of...umm...a challenge...to answer the questions because you're kind of like... uummm...trying to be fair, but you're also trying to be truthful and...uumm...but without, you know...uummm...damaging." Sara's comment also shows that students are likely not aiming at trying to hurt the instructor with their reactions to the class, especially those like Sara who wrestle with how to be truthful and fair but without saying something hurtful. Like Dagmar and Margaret, writing comments has a personal connection for Sara as well. Sara likes being able to add her comments because it gives her ratings "a voice."

In his opening remarks, John described the value of ratings for him. "Basically, I think it has a potential of being a good communication device between...umm...a student and an instructor or the administration of the school so that you let people know what you've appreciated and...also give feedback to things that kind of need to be corrected or should possibly be corrected."

Greg actually discussed having a higher evaluation standard for classes in subjects that he is interested in as opposed to general education required courses. For him, his higher expectations for courses he wants to take cause him to take a greater interest in rating those classes. He explained it this way:

In courses that I don't care about where...it's...a little easier for me to skate through and get an okay grade...um...as long as the teacher isn't like grossly...incompetent or something, then I'll rate them...pretty...good...okay maybe...pretty okay...unless they're exceptional teachers. Then, you know, I'll rate them exceptionally, or if they're terrible, I'll rate them terribly. [For the classes I am interested in] I judge them...uh...I make harder judgments because I want to know more. And if they aren't providing it, or...or if it's very slow or if they, they don't...seem to, uh...develop...like the course material or their ideas fully, then I, I get a little...a little annoyed if it's actually something that I want to know about.

It is significant to consider Greg's remarks in the context that he is also the participant, as discussed earlier, who does not write comments for those instructors that he feels are "adequate." Without an interest in the course, he has no interest in writing comments. "It's more like whether I, I even care about rating the instructor [sic], the instructor..." Although he claimed he withholds his comments if he thinks the instructor will "take it too harshly" or if he feels that his comments will not be anonymous, he also asserts, "I don't...purposely, uh...write down wrong things.... I try to be fair in what I, what I say...and keep in mind how, where they seem to be going and what they seem to be doing...and try to fit my comments into that." Therefore, even with a student such as Greg who will withhold comments and is just as likely to complete only the Likert-scale questions and leave, he claimed to be honest and fair when he does write comments.

For Jaime, the comments section provides a place to praise the instructor. "I just want to let them know they, like, thank you for what I've learned from them so far, or...that I think

they're doing a good job. I don't know if that's...ridiculous, but I wan- I feel like I need to let the person that was reading this exactly why I write what I write...or how much like them or what maybe they could do to change a little bit or to get through to some students more."

John related having similar feelings when it comes to criticizing those instructors that he finds to be effective overall:

It's easier to tell the teachers that are effective the good things about them and letting them know what may not be as effective...at least from my personal experiences. But...uh... not wanting to come across as attacking them in any way so...uh...but I guess I would probably lighten it up to...indicate that I was more satisfied than dissatisfied whereas somebody that I was kind of...aaah, even keel, the, their, they were okay, they weren't great, here's some things...it's, like, okay, well, I, I like this, I this, I didn't this, I didn't like that, and I'd probably be more blunt about what was effective and what wasn't.

Both Jamie and John indicate that while it may be more difficult to provide negative criticism of instructors who are effective overall, they will still do so. Moreover, like Sara, they will try to couch their criticisms in such a way so as to avoid hurting the instructor's feelings, which appears to be the motive behind trying to "lighten it up."

Two of the participants even indicated that they have a great respect for all instructors that likely accompanies them into all their classes and, if anything, makes them reluctant to be too critical. Andrew demonstrated this reluctance when he said, "I think it's like teachers...they're teaching, you know. It's a...it's a...honorable profession... and that...I don't know...I, you shouldn't really...you're a student. Your job is to learn from the teacher, not to pick apart the teacher what they're doing wrong and stuff." Annie also expressed a similar kind of respect for instructors. As a student who tried college out of high school but then dropped out and is now continuing her education in her 30's, Annie has a new respect

for instructors:

I take a lot of things into consideration. I take in...um...I, I...um...uh, respect for, um...for, um...the degree that they hold because I realize how, uh, I mean the amount of schooling I have had is nowhere in comparison to theirs, and it's hard just staying, you know, I mean, just...um...staying in school. It's a lot of work, you know, and so I have this, like, you know, they're almost like demigods to me now.

Trying to be objective even extends to instructors that students may not like personally and regardless of the grade the student expected to receive. Susan remarked, "I don't think I would ever...use a dislike of a professor as a person as a way to give them a bad evaluation if they're actually effective. I don't think I would ever...I would hope I'd be objective enough to not do that." Like Susan, Asia would try to be objective despite whatever grade she thinks she is earning. As she put it, "I've always been true, and I've always been fair, and I've always put down what I felt...regardless of what grade I was getting at the time...you know. And I, and I do try to do the comments because the comments are good. It's good feedback." Asia does not stop with herself. Later in her interview, she expressed that this type of fairness and honesty should extend to all students:

I believe that...all students should be fair and really put down what they feel...and don't over...evaluate a teacher because you like them...and you think they're going to give you an "A" or whatever or don't under-, uh, evaluate a person because you're getting a bad grade 'cause that...that doesn't...really affect, you know, that, that...she's doing or he's doing his or her job. You're just not putting in the work.

Overall, as the comments indicate, the participants I interviewed generally take it as their responsibility to be objective, fair, and honest in completing SRTs. Yet these are the same students who also acknowledged that they withhold their comments or overrate certain instructors when, as I content, they find themselves in a climate of fear. What appears to

account for their equal willingness to be candid and sincere is their being in a climate of trust.

Trust – The Key Factor Underlying Candor and Sincerity on SRTs

The participants did not speak directly about feeling trust, yet in describing conditions that affect how they respond on their ratings of instructors and classes, they presented a number of factors or conditions that create an environment in which they feel safe when providing comments. What emerges from these elements, then, as a whole is what I label as a climate of trust. I have grouped these factors into three general categories: preconceptions about their responsibilities in completing ratings; knowledge of, beliefs about, and attitudes toward the rating process; and student perceptions of instructors before and during courses.

Preconceptions about Responsibilities and Factors That May Impact Ratings.

Having a sense of responsibility comes across in the remarks of Dagmar, Margaret, and Sara. Students also probably share Susan's basic sense of responsibility about at least not intentionally lying on a rating. "I would probably...leave them blank. I don't think I would *lie* and say, 'Oh, this teacher's wonderful, great,' you know, brown-nosing kind of thing." In Susan's case, she would withhold comments rather than lie. Other comments from various participants demonstrate that they have other preconceptions about what their role in the process is.

A sense of responsibility also appears in their beliefs about how they and other students should or do respond when writing comments. Asia discussed the approach that she takes and that she thinks all students should take when writing comments:

You just have to come it towards, you know, cross it the right way. You can't come say you're a bad teacher and this and that and the other. Come across like an adult and say this is how I feel about your class. I love your class and...everything

is great except for the workload is just intense. What can I do...to...accommodate...or juggle all these things and still produce good work? They might have suggestions. They might not take it the wrong way.

The key phrase is “come across like an adult.” In other words students should approach the process maturely. Margaret’s comment about not simply criticizing an instructor for wearing a pink dress or not merely saying that “the class sucks” made the same point that Asia was making. Also, Susan’s discussion of trying to avoid allowing her personal feelings about an instructor from affecting her objectivity provides another example of the sense of responsibility that students have about completing ratings and shows a student acting in a mature manner.

The same level of maturity concerning the importance of truthfulness appears in something that Sara said. “The effects of lying or not being truthful or not having creative expression is [sic] totally detrimental to the individual as well as to the collective whole.” Sara comes into the class with a preconception of the impact that not being truthful can have at a level that extends beyond the individual and perhaps even beyond the classroom. The importance of honesty appeared again in one of Sara’s later comments that echoes Margaret’s. “And honest doesn’t mean being a jerk, you know, just be like, you know, I really hate the way you look today. I mean, honest means honest is about...being strategically honest like so saying, ‘You know what you are, who you are, and what you’re trying to be and do, you know. How’re you going to do it?’ And we’re trying to find a way that makes that work.” These comments from Asia, Margaret, and Sara reflect basically the same general attitude about their responsibility in the process not only to be honest but also to present constructive suggestions in a mature manner.

Another way in which student preconceptions manifested themselves in the interviews occurred when Susan and Greg referred to what are, in effect, norms that they use in the rating process. The application of these norms demonstrates a level of sophistication as well as the thought that at least some participants likely bring to the process. Susan has developed her norm based on her collective experience with instructors. As she put it, “I would probably compare one instructor to the broad base of instructors I’ve already...had because you ha-, most of the time you have to give them a numerical rating, so you have to have some sort of basis for...average.” Greg has also established a norm for rating instructors but based on a different criterion. “I think about...what...I want to get out of the teacher, or at the very least, what I think they [sic] should be providing. And then I measure that against what they actually are doing.” All the participants expressed factors that they consider when completing ratings, factors that I will discuss later in the context of instructor behavior. To an extent, these factors reflect criteria for the norm that they use.

That students are discriminating in how they rate instructors occurs both during the ratings and outside of class. As mentioned earlier, Asia distinguishes between “hard” and “lenient” instructors.

Like I tell a lot of people...you know, this teacher if they want to know a reputation of a teacher, I go, well...this teacher is good. You *will* learn. You *will* work hard. Um, so if you, if you don’t want to put out good work and put the effort in the class, then maybe you shouldn’t take that instructor. So if you want to *learn* and you want to bust your *butt* [laughs], you will work hard that semester, then take that teacher.... Yeah...so some tea-, so some people say, you know, I really want to learn math or really want to know this algebra, then take this professor. But if you want to just *get by*...and you’re really just, you know...if she’s very lenient or he’s very lenient, then take that teacher.

Clearly, these comments indicate that Asia has a preconception of what constitutes a “good”

instructor: someone from whom she will “learn” even if it means working hard. When Asia takes this preconception into a class, it indicates that she will be using learning as a primary criterion in rating instructors and, again, illustrates how seriously some students take their responsibility in rating instructors. To an extent, students’ preconceptions appear also in their views about the rating process.

Knowledge of, Beliefs, and Attitudes toward the Rating Process.

I already discussed in a previous section about how misunderstandings about the rating process can lead to students believing either that the faculty and/or administration will react negatively or that no one may be paying any attention to the ratings and the subsequent student responses. In contrast to these beliefs are comments from participants who indicated having more trust in the process. John exhibited a realistic view of the function of the rating process and the need for providing some balance in the ratings. He also showed an awareness of how administrators would likely react to extremely negative comments about an instructor:

...[S]omebody might go in there and totally slam the teacher, and, okay, that might flag an administrator to look at something. But, you know, they’re out to get that teacher, and it’s really evident in their paper. So an administrator’s kind of going to say, ‘Okay, well what happened to want the student to get that far.’ But it’s kind of going to discredit a lot of the things that they say in there because it’s such an attack whereas...you know, some people are, can take a review and say, ‘Okay, these were the highlights. These are good things about the class. These are the things that I would have done differently or I think could be improved’ and then somebody’s got a really diverse understanding of it.

Participants also expressed positive attitudes toward the rating process. “I think faculty evaluations are essential,” said Margaret. After this comment, she added that it is important to be able to give an instructor constructive comments when problems occur and provided an

example. “I felt that she had a lot of knowledge, and I could really say, you know, ‘Maybe if you notice when people are asking a million questions about a point, it’s not that they’re stupid or they didn’t read it. Maybe they don’t understand it.’” Later during her interview, Margaret also provided her view about the value of having multiple questions on the ratings form, targeting a variety of aspects of teaching:

Because of the way questions are worded in most evaluation forms, it doesn’t feel like you’re saying the teacher is a bad person or even a bad teacher necessarily. You’re just saying, ‘Well, in this area blah, blah, blah.’ They need to improve this, and that’s when I think the real value is about teacher evaluations because they’re not a one-question thing: do you like this teacher, or do you like this course? That’s not there; there’s so much more detail, and that’s the real value of them is that it gets students to think about specific things.

Remarks such as Margaret’s provide evidence that students see value in the rating process. In fact, Annie expressed disappointment at not having the opportunity to provide ratings for some adjunct instructors that she felt were not effective teachers. “I was very concerned about these teachers, and I, um, was wondering when those questionnaires were going to come in, but they never did.” Also, according to Sara, being able to add comments creates value for her as it allows her to give “a voice” to her ratings, and “it sort of draws a picture, you know, that’s more descriptive and specific and, um, you know, less vague.”

Knowing the purpose of the ratings can impact student attitudes toward providing candid comments. Andrew, the participant who feared that the administration might cancel a course and/or deny students credit, said that he would take a completely different approach toward writing comments if he knew that the purpose of the ratings is to improve teaching. “If I knew there was just so the teacher could improve, then I’d be a lot, yeah, then I’d, then I think I’d be perfectly honest.” Besides student knowledge of and attitudes toward ratings,

preconceptions and/or expectations can influence student trust.

Student Perceptions of Instructors.

In the next section of this chapter, I will discuss instructor behavior that appears to contribute to establishing a climate of trust in the classroom. The way instructors relate to students will influence how they view the class. This subsection closely relates to the next section. The distinguishing characteristic is that some of the participants spoke about their positive attitudes toward instructors separate from behaviors that instructors display.

Annie holds most instructors in such high esteem generally that she thinks of them as “demigods.” As discussed earlier, she is an older student who began college right out of high school, dropped out, is a single mother, and has returned to complete her college education in her early 30’s. Now realizing not only the value of an education but also the effort that goes into doing well in college, she has a newfound respect for her professors in general. Therefore, when instructors are knowledgeable, they have already earned Annie’s trust. Recall that Annie is one who overrated her instructor because she did not want anything negative to happen to him. His being kind and nice likely reinforced the positive attitude toward instructors that she apparently takes with her into every new course. However, Annie also indicated, as previously mentioned, that she would have liked the chance to provide negative ratings for some of the ineffective adjunct instructors from whom she has taken courses;

I’ve had crappy teachers, and I’ve never had questionnaires come in with those teachers.... Part-time faculty, and they were crap, in my opinion...unfair, inexperienced, and graded very unfairly and didn’t really present the material in a way that, uh, y-you know, the majority of the class could understand and, um, conceptualize an-and bring it back and do well on a test.

So being predisposed to view instructors as demigods does not guarantee that they cannot fall from grace if they prove to be “crappy teachers.”

Also discussed earlier, Andrew also views teaching as “an honorable profession,” so students should not “pick apart the teacher what they’re doing wrong and stuff.” Even Sara made a comment to indicate she, too, holds instructors in a higher regard. In trying to recall some of the questions on the rating form, she said, “I don’t remember the questions being about the, the curriculum per se...um...but then again, you know, the students aren’t experts on the curriculum anyways.” These various remarks indicate that at least some students have a high enough regard for instructors and their expertise that they likely enter all classes with a degree of trust. Whether that trust remains will depend on what the instructor actually demonstrates in class.

Besides having a prior respect for the instructor’s status as a teacher, some participants related liking the instructor. Jamie admitted that “how much I like the teacher so far” is a factor she considers when completing ratings. In discussing the instructor she overrated, Annie prefaced her remarks with, “I had a teacher. I just really liked him.” Both participants made it clear that liking the instructor was not the only factor they consider. Jamie, for example, added liking the instructor at the end of list of other factors related to instruction. Whether students give instructors they like higher ratings simply because they like them or whether students like the instructors as a result of their being effective at helping them learn is unclear. Regardless of the reason, when students like an instructor, that signals a degree of trust.

As discussed above, preconceptions, knowledge of and attitudes toward the rating

process, and student perceptions of instructors are factors that emerged from the interviews that reflect positive views of instructors and the rating process. These factors reflect a climate of trust that, to a great extent, emanates from positive instructor behavior in the classroom.

Positive Instructor Behavior and the Culture of Trust

Participants described a number of actions that instructors and administrators have taken or could take that would encourage candor and sincerity in student responses on ratings. Additionally, they cite behaviors and characteristics that create, in effect, a climate of trust. The same categories for behaviors used to present the findings in relation to behaviors that lead to a culture of fear provide a parallel structure for the behaviors that contribute to creating a culture of trust: actions related specifically to the ratings, communication skills, attitude/personality traits, treatment of students, and pedagogical skills and practices.

Actions in Relation to the Ratings.

Perhaps the most significant action that instructors could take would be to encourage students to be candid even if some choose not to take the instructor up on the offer. If the instructor comes across as sincere in the request, the instructor may meet with some success in obtaining honest responses. Such a student reaction would be true for Margaret, who said, “I have had instructors say, you know, when you give this feedback, really think about what you’re going to say because I can benefit from it, and that, those are the ones where I just really...lead just pours out of my pencil.” Jamie did not relate any personal experience with an instructor who encouraged honesty; however, in responding to the question about what

instructors and/or administrators can do to promote more candor and sincerity, she suggested that instructors could encourage students to be honest. "I don't know, maybe they could just let everyone know... be honest on this. It's only going to help me. It's only going to help the other teachers.... maybe just letting...to encourage them. That there's issues with something I'm doing or if you think there is something I should be doing something different, feel free to write it." However, responses from other participants indicate that it will likely take more than just encouragement of honesty to convince students to act on that request.

Like Jaime, Diane also suggested that instructors encourage student honesty by "maybe talking before...the...um...evaluation's handed out...what we're trying to gain from it and what you're not trying to gain from it and...emphasizing the fact that they want us to be sincere in, uh...in what we're writing down." But Diane took her suggestion a step further, acknowledging that it would help to know that the instructor will act on the comments to affect changes. "I think they [instructors] could...make it seem...um, that, it really...it really matters as to what you put down, like that something will change if...we dislike it. You know, if enough people dislike a certain thing that it will be changed versus like...just keeping it the same way and just fill out the form, you know." This point was important enough to Diane that later in her interview she reemphasized it. "I think just as long as they...um...give us the feel that we're going...to change something if...it's being disliked by a large number of students, then I think that's, that's the best or the most that anybody could ask for...with the evaluations...as long as something was going to get changed if a lot of people didn't like it." If instructors were to follow Diane's suggestion, they would, in effect, give students a sense that the instructors value students and their comments. It stands to reason that when people feel as if someone values their views, they will be more forthcoming

and forthright in giving them. Andrew actually provided an example of an instructor who did exactly this.

Although the following description Andrew provided is somewhat lengthy, it is important to quote in its entirety to demonstrate that it takes more than just encouragement to promote candor and sincerity:

Yeah, I think I'm more candid when the teacher actually...is like this is just for me, you know. I had a science teacher in high school who did that, and I, I respected him a lot, you know. And he wanted you to be honest, you know. And, um...yeah, I put things, you know, like maybe...this weekly quiz or something could be structured differently, or this is kind of, you know, stuff like that...like...course material and stuff. And then he had a really good class, too. He's, I think, the only teacher I've had who did that. And his class was well structured, and it's like...it's like he actually paid attention to what people were talking about and tried to improve in the areas where people wanted... 'cause like he would start each day's classes, science class, you'd start out five minutes, he'd make, uh...on an overhead projector, he would have, uh, like, news articles related to science, different things, so it was like, oh, this is like, I guess people would say like, you know, we have to relate this to real life or something, so he would show how it was related to real life. I thought that was good.

Besides the encouragement from the instructor to be honest, Andrew's instructor showed students that he was serious about obtaining their comments by acting on them. His bringing in news articles to help relate science to real life came as a response to student comments. When instructors prove to the class that they are using student comments to try to improve their teaching—prove that asking for comments is not merely empty rhetoric—students will likely be more inclined to trust them enough to be honest. Andrew characterized his instructor this way: "It's like, oh, the teach-, the instructor's trying, you know, like, he's trying to improve, you know. He's not just stuck in his...how he's doing things. He's trying to improve. He's, he's open-minded. He's open to suggestion." From Andrew's description, it is clear that this instructor elicited comments early during the class so that he could act on

the suggestions as appropriate.

That the instructor must begin to create a climate that allows students to be honest long before students complete a rating appears in Sara's remarks as well. "I think basically it's like the surveys only serve a purpose if the student already has permission looong ago, not just at that moment when they're giving the survey." Instructors who encourage students by creating an environment that makes it safe for students to comment will likely elicit candid and sincere comments on their ratings.

Related to the suggestion that the instructor act on students' comments is that recommending that students need to understand the process and how instructors and/or administrators use the ratings. As discussed earlier, several participants acknowledged a general uncertainty about whom, if anyone, looks at the ratings and what, if anything, happens as a result. Diane recommended that students be informed, "maybe make the students more aware of the process of the evaluation goes through after it's been handed to us." Shortly after she made that comment, she added, "I think that...uh...that would definitely be encouraging if you actually had a, um...the ratings and the statistics and everything that went along with evaluations. I think that would be encouraging...at least to have some kind of follow up basis as to what happened...with that piece of paper that you got handed to [sic]." During her interview, Annie included a similar recommendation: "I think it would be good to, kind of, give, um, an overview of exactly what these questionnaires are used for."

Susan also suggested that it would help for students to know the process but for a different reason. "I think there's [sic] always going to be kids who really don't like

instructors and they're going to have things to improve upon and maybe to know that nothing negative is going to happen to them just because they have a bad...evaluation, like maybe knowing the process...of, you know, just because they have a bad eval-, evaluations, it means something bad's going to happen...maybe they get talked to." In this case, Susan appears to have a concern about severe consequences for an instructor (the implication being disciplinary action or non-renewal) but for dual reasons: either comments that may legitimately point out areas the instructor needs to improve or intentional underrating from disgruntled students. The belief is that with a better understanding of the process and how the instructor and the administration use ratings, students may feel more comfortable about providing comments and being sincere.

Besides recommending that the institution provide students with information to help them understand the process and the use of the results, some of the participants also offered suggestions about how instructors and administrators conduct the process itself. The current practice at the institution that the participants attend is for peers to conduct classroom observations and then to administer student ratings at the end of the class that they observe, during which time the instructor leaves the room. John suggested that it may help to provide students with advance notice of when the student ratings will occur

...so that students can take some, those who want to can think about it and take some time to either make mental notes or make notes to themselves about the highlights and the not so great moments of the classes so that they can be more objective over the whole class rather than dealing with the things that are on the top of their mind that happened recently.

Perhaps if students had time to consider how they can phrase their concerns in such a way to avoid being connected with the comments, they would be more willing to express them.

Fear of being identified, as presented earlier, is one of the most common fears that results in a lack of candor or even a refusal on the part of some students to write any comments at all. Even if instructors do not ask students to be honest with their ratings, ensuring anonymity is essential to obtaining at least some true responses from any students who have any reluctance about being candid. Without repeating the previous discussion, I also discovered that some of the participants offered specific suggestions about ways that the instructor and the administration could help to ensure anonymity. John provided another idea to address the concern about anonymity. “The other thing that I think might be a good way of doing it is if the evaluations were done after the end of...after the final grades were published so that there’s no looming fear of, ‘Oh, if I don’t say certain things my evalu...er, my grade might get dropped, or maybe if I’m really nice my grade will go up’ or whatever.” Andrew also recommended, “Take them at the end of the course as opposed to in the middle of the course, you know, in its evaluations.” Their idea is not new. Some institutions make a common practice of waiting until the end of the course to conduct student ratings. What is interesting is John’s thought that using this practice would address not only the issue of candor and its relationship to fear of grade retaliation but also the issue of overrating an instructor in hopes of obtaining a better grade, which was Dagmar’s concern. Later, John even suggested giving students the ratings to take home to complete and then mail in; however, he conceded that this might not be the best method because “a majority of the students wouldn’t care, and it would get...either they wouldn’t care or things would get busy and they would forget to do it, and there’d be much less feedback because of that.” However, John was not alone with the suggestion of mailing the ratings.

Susan also indicated mailing the ratings might help establish trust. “So I think having

that trust of knowing...so either being able to mail them in or, or to have someone sit, different in the room at the time, maybe two professors at the same time swap to do it or something like that to assure that it's going to be anonymous, I think is really important."

Related to actions are an instructor's communication skills.

Communication Skills.

Lack of communication or instructors' basically closing off communication through their behavior in class, as discussed earlier, contributes to creating a climate of fear for some students. Participants also emphasized how much they value good communication with instructors. For example, in listing factors that Asia considers when rating instructors, she included communication in her list. "You know, how well she works with others...if she's...I can present her, I can talk to her or him."

For Dagmar, communication, or "dialoguing" as she put it, comprises a critical ingredient of creating a climate that will encourage students to be candid and sincere when completing SRTs. This process begins when students enter the class, not just when students have to complete their ratings of the class. "From the first days students come to give them really a picture of, uh, or an open invitation that administration and faculty are in a dialogue with the students and not in a monologue. So the students is [sic] not in a position of, uh, being told what to do or not to do rather than being in a dialogue how can we do something together, and that's what education is about. It's in a dialogue."

The concept of dialoguing is so central in Dagmar's view that she discussed it in the context of the entire college. She began by emphasizing that faculty members need to learn how to engage in it with their students. "I'd say a faculty need to do some sort of training as

well...towards how to dialogue with the students because it's sort of, you know, it works its way downwards." The inability to communicate well she saw as pervading the institution as a whole. "What's happens on a bigger scale happens on a small scale. So my...suspicion is that faculty, as well, doesn't [sic] really know how to communicate, how to dialogue with administration or to really bring forward what they need, what they want in order to do...to be an excellent educator." From there, she described the various levels of fear that extend in a "ripple effect" from faculty to administration, from administration to the Board of Trustees, and from the Board of Trustees to the state—what I discussed earlier as becoming a culture of fear.

In effect, Sara also describes a culture that involves communication when she described the importance of practicing and encouraging critical thinking throughout the institution. From her perspective, not only should the instructor be engaging students in critical thinking in the classroom, but administrators need to do the same with faculty and staff. "The administrators who know this, they need to practice it with the instructors who don't know it or who do know it. It has to be continual, a consistent, you know, process, and, um, then you're talking about then overlapping to the staff and, I mean like, student staffs as well as, you know, fulltime employees and, and then anybody else who's in the sector because, you know, it's not a vacuum, you know." Although Sara's point relates more to the significance of critical thinking, it is through communication that people engage in critical thinking—instructors asking questions and students responding. It seems significant that, like Dagmar, she envisions the process needing to occur throughout the institution, not just isolated within the classroom.

As important as openly expressing one's views is listening. Communication involves both speaking and listening. Miles emphasized the significance of listening when he described the instructor who essentially refused to listen to student questions or concerns when students complained that they could not hear him when he faced the board. According to Miles, listening to students is about all that administrators can do. "I think the only thing an administrator can do would just be to be encouraging as far as the...uh, uh...listening to questions and listening to different things that students have to offer." The importance of communication—both speaking and listening, engaging in dialogue—emerges as a key element of promoting a climate of trust in the classroom, a culture of trust in the college.

As the previous statements about communication indicate, particularly Dagmar's, when communication begins in the first days of class, not merely at the time students complete the ratings, the likelihood of establishing a level of trust within the classroom increases. Promoting communication from the outset of a course provides one means of gauging an instructor's attitude toward students. From the comments of the participants came other attitude and personality traits that had a positive impact on them.

Attitude and Personality Traits.

The attitude and personality traits that participants mentioned, as well as comments related to the treatment of students (discussed in the next section), came mostly in response to the first question about factors they consider when rating instructors. These traits, when positive, would arguably have an impact on the climate that the instructor establishes in the classroom. For instance, Margaret expressed that she considers how the instructor interacts with the students:

I think how the instructor interacted with me and my understanding of the topic, and I also think about what I observed in the classroom. The instructor interacting with other students and what they seem to not assume what they say but how...how I've observed moments happening in the classroom.

This interaction will demonstrate the instructor's attitude toward students. Diane began her interview with this statement: "I consider the teacher's personality and her willingness and, um, enthusiasm she presents to the class...and her, um...her knowledge about the subject and how she makes the students either excited or not excited about the topic that we're about to study."

Sara also mentioned enthusiasm for teaching, distinguishing between instructors who are simply teaching the curriculum and those who have a passion for what they are teaching. "It's like you see teachers and they're teaching the curriculum. They're teaching from the text, and that's great. But there's also this like other...you know...level of teaching like if the teacher really loves what their doing, then they're like teaching joy." It is possible, too, that instructor enthusiasm and passion contribute to a factor that Diane considers, "...how much fun the class is, but that's really depending on the teacher and her attitude and her...um...how she feels about class participation and sections." As Diane's comments indicate, effective teaching likely combines both attitude and active learning that engages students.

In discussing actions that faculty members could take that would promote candor and sincerity, Miles added other characteristics that have a positive impact on him. Of importance to him is "the way they [instructors] come across. Uh...if it's...if they're real..." Miles provided a list of traits that appear to define what he meant by "real": "But basically, uh, sincerity...uh...honesty...uh...diligence...those are the things that I look for when I'm,

when I'm rating a teacher." Later in his interview, he returned to the theme of being honest with students as a practice that instructors could engage in that would help produce more candid student responses. "Be as honest as possible. I mean, um, um, sometimes being vulnerable, admitting some of your flaws, you know.... You know...making yourself...you know, making yourself vulnerable I think helps a lot 'cause then you put yourself in the students' boat, you know. You're right there with them." Miles, an older student who is also a musician and has taught music, used his own experience as an example of making oneself vulnerable. He related how on one occasion he admitted to students that he had reassessed what he had told them in the previous class and realized that he was not giving them what he really wanted them to have. In other words, he admitted to making a mistake, an action that likely helped establish more trust with his students.

Students are likely discriminating, however, when it comes to displaying vulnerability. They may view an instructor who admits that something did not work the way she or he had planned or who even admits a mistake as being human, but they may view an instructor who admits being unprepared as incompetent as Sara's comments indicate. She described what an instructor told her class on the first day of class:

This is the first time I'm teaching the class...so I don't have a syllabus...and I don't know what I'm going to do, and I don't know how to grade this because this is sort of, uh, you know, this is so different from, from what I'm used to, and it's sort of, um, you know, um... you know, how do you grade art...sort of...you know, kind of thing. Um...so...on the one hand, you know, I, I did feel...it was an excuse not to be the best....

Closely related to the attitudes and personality traits that instructors display in class is how instructors treat students. Just as actions in class such as failing to listen to their questions can produce a climate of fear, positive actions can create a climate of trust.

Treatment of Students.

As Annie related (cited earlier), she overrated an instructor who, despite being disorganized, was “very kind, very nice.” The way he treated her caused her to say that she “just really liked him.” The implication is that when instructors are kind, they signal that they care about the students. In turn, some students will try to protect them from harm by overrating the instructor (Annie), perhaps withholding any critical remarks, or by “toning down” the comments (Andrew). If students feel safe about the process, they are more likely to include comments that suggest improvement that instructors can make.

The terms that John used as qualities what he looks for in instructors are being “student friendly and learning oriented.” For John these characteristics appear to result in a certain treatment of students and translate into his liking the instructor. During his interview, he distinguished “effective” instructors from ineffective ones. The implication is that he likes effective instructors, not necessarily because they are kind and nice, as in Annie’s case, rather because they are effective at promoting student learning. Consequently, it affects how he presents his suggestions for improvement as opposed to how he would present suggestions for less effective instructors. John put it this way:

I would probably phrase the teacher that I thought was really effective a little bit lighter...so that...to make...I guess I would be afraid that if they saw the criticism and it wasn’t worded lightly, like, here it, well, you know, I really enjoyed your class and blah, blah, blah, and fluff that part up, and...um...kind of make the criticism kind of like an insignificant thing, and, you know, I really enjoyed all this, but there was kind of this one little thing that I feel could have been done differently whereas somebody that I didn’t feel was as effective I’d probably would be more blunt in my answers to them.

Other participants noted other ways in which instructors treat students that may also

contribute to students' liking these instructors although the participants did not specifically mention this connection. Jamie, for example, pays attention to how equitably instructors treat students: "I watch the equa...like, I don't know if it's equality, but how the teacher...treats everyone in the class." Asia expressed essentially the same idea; however, the term she used to characterize how instructors treat students is "fairness." She also added "how well she interacts with the students" to her list of factors that she considers when she completes an SRT.

Sara described a condition that will lead to either higher ratings or, at least, honest ratings that results from instructors freeing students to be honest. Although she does not mention specific treatment that would cause students to feel free to be honest, it seems clear that how the instructor treats students in class is a critical component. She described the conditions this way:

Okay...um...if the student feels that they are liberated...in this educational environment...that they had already felt that they can be honest [half laugh] about what they're feeling or about what they're...I'm not talking about random information. I'm talking about the, the class...their education. So if they're, if that's already been established, when the question...when they're looking at the question, they're going to say excellent or they're going to at least have the honesty to say poor or whatever their answer wants to be because they know that they can do it. They've already been given permission.

Sara returned to the theme of being given "permission" to be honest later in her interview. She contrasted an "oppressive" environment in which students are "reprimanded or punished" (as discussed earlier) with an environment that promotes critical thinking and openness. "It has to be...um... okay to be honest...anyway, not even around the surveys. It has to be, you know...uh...okay. You have to be liberated." From her perspective, being honest on SRTs begins long before students have the opportunity to rate the instructor or the

class. Simply to urge students to be honest at the time they receive the SRTs is not likely to cause Sara to be honest necessarily. As she said, making students feel as if they have permission to be honest, that they are liberated, begins early on in the class. She also said, “It’s a practice of being open. So if you say, ‘We need to create an atmosphere of being open,’ then you’ve, you missed the whole bo, the boat.”

Giving students permission to be honest is one means of giving students respect. Another way of treating students that demonstrates respect for them is by grading and returning their work in a timely manner. As cited earlier, Diane commented on the fact that she finds that sometimes instructors will assign more work than they are willing to grade and that when they neglect returning the work within a week, it demonstrates that they are “not as dedicated to correcting the work as [they] are to assigning it.” Returning student work in a timely manner reflects how instructors treat students, but it also relates to the final area that appears to impact the creation of a climate of trust—pedagogical skills and practices.

Pedagogical Skills and Practices.

The interview questions did not aim directly at asking the participants to discuss the pedagogical skills and/or practices that may influence how they rate instructors. Therefore, the comments that did emerge that relate to skills and practices do not necessarily reflect what the participants might have said if I had specifically asked about these factors. However, participants did mention elements of pedagogy in terms of factors that they consider when completing SRTs.

The pedagogical practice that students mentioned most frequently is organization. The first sentence that Jamie uttered in response to the first question concerned organization. “I

think about...the teacher's organization of teaching....what their course outline is and how they stick to that, how often I get the syllabus, is that what we're going to follow and what I can expect to learn...throughout the semester." Sara's example of the instructor who had no syllabus demonstrates the importance of having an organized plan for presenting the course. Sara was also the participant who recommended a course in how to write a syllabus for instructors. Annie commented that instructors should "stick to their syllabi." In listing factors that she considers when rating instructors, Asia included "organized and presenting the structure." She also listed "how prepared she is with her daily, uh, assignments" as a factor and at another point in her interview, she added, "It just really helps when the teacher is prepared...presenting the course and not leaving any...things out." Annie also noticed her instructor's disorganization. "The way he conducted the class and his...the way he lectured was a little unorganized...um...in the sense that, um, it didn't really follow the chapters we were reading. You kind of go off on tangents." However, because she liked him, she overrated him. The contrast between Sara's and Annie's reactions to their respective instructors illustrates the complexity of ratings. In Sara's case, she interpreted the instructor's disorganization and lack of a syllabus as a sign of not caring. As for Annie, she willingly overlooked the disorganization because the instructor was so kind and nice.

Participants also mentioned instructors who are effective in helping students learn. According to Asia, she considers how well an instructor prepares students for what they will be tested on. "I really look at an instructor and the subject...the subject and if she, if she put, or he, presenting it fairly and preparing us for, um, the next level or exams." Later in her interview, she returned to this point about instructors ensuring that students are learning. "I look at that because that tells me that, you know, you're really taking the time with your

student, not individually, but taking the time to just break it down even though we are on a time schedule as far as the class and everything. But you need to prepare your students and na-, and not have any...uh, surprises.”

John described his candidate for outstanding instructor as someone who demonstrates his concern for student learning by using a variety of learning techniques to teach to different learning styles:

I’m amazed with his dedication to teaching, and he will take your tests and whatever you failed on the test, any question you got wrong, he wants you to learn that material. Obviously, you didn’t make it on the test, but you still need to know this. It’s important. So go and write paragraphs, and each paragraph is supposed to be detailed of why...what you put was wrong and why what the correct answer...why is it right, so that it’s really start drilling it into your head again on a whole different level of what’s going on. And he takes the time to do, like, theater in class so that people can get a visual and act out what’s going on in the processes, does stuff on the chalkboard, does stuff with lectures, does stuff with PowerPoint. So he’s incorporating every type of learning style into his lectures. So maybe somebody’s not...doesn’t like what he’s doing, but I think encouraging his methods are, would be a good thing because I think he can reach out and grab everybody and bring them in and help have them learn something whereas a lot of teachers are very focused on how handle learning styles, and though they’re good teachers, they’re only good for specific learning styles. And not everybody can benefit from that.

Such an instructor is bound to create a learner-centered classroom that would undoubtedly create a climate of trust. Andrew’s example, described earlier, of the science teacher who acted on the comments he received from students to improve student learning provides another example of an instructor who demonstrates a commitment to student learning.

Sara’s ideal instructor is one who encourages critical thinking in the classroom. She contended that instructors need to teach students that critical thinking is part of “their job”

but that instructors fail to do this. “They don’t...they’re not taking the time to make sure that, you know, students are understanding what their job is,” which she then proceeded to describe as “teaching students how to critical think in every class, you know, not just in Critical Thinking [the college has a class called Critical Thinking]. Um, you know, there’s a...a...a teaching of ethics, you know, so that their emotional integrity is in place.” Sara returned to the topic of critical thinking later during her interview, this time linking it with organization and planning:

Question the students. You know, um...probe them, provoke them to think...critically. And, um...you know, there definitely has to be, you know, a, a strategic focus. It can’t be, it can’t be so loose that there’s, you know, random knowledge being thrown around, you know...um...or irrelevant issues brought up, you know. Um, that happens, and you lose the class...solidarity, you know. Um, so there has to be a, you know, a strategic plan, you know, from the beginning.

Teaching students to question critically but in an organized fashion is critical, according to Sara, to ensure that when it comes time to complete ratings, students will respond candidly and sincerely. Critical thinking is part of the process for making it clear to students that they have “permission to be honest” to use Sara’s expression:

And so, you know, um, there, there has to be an environment, uumm, that is, you know, um, just...has to be a practicing, you know, environment. So it’s like, you know, the instructors really need to question their students so that they practice, so that, that they teach...so they teach like almost subconsciously to question the instructor.

In other words, Sara is advocating for a classroom in which the instructor promotes an environment that requires students to practice critical thinking routinely in everything they do. Creating an environment of critical thinking is a key element of fostering a climate of trust if it does provide students, as Sara contended, with a feeling that they have permission

to be candid and sincere. Without the promotion of critical thinking in the classroom, Sara believed that the point of education changes. “So it’s like if we’re not being taught to think critically and it’s not okay to challenge instructors, then really those two critical points, there’s really no, uh, education in the fundamental sense of, you know, growth and enrichment. We’re just learning to, to, uh, get a grade or to get money or whatever the material, you know, uh, end is.” When instructors teach students to think critically and encourage them to challenge what they are teaching, one could argue that these instructors are, in effect, sharing authority with the students, treating their views with respect and causing them to feel valued. It is difficult to imagine that such a classroom atmosphere could not help but create a climate of trust among the students and lead to candor and sincerity when completing SRTs.

As a result of being in a class in which the instructor encourages critical thinking, Sara sees the rating process as taking a new direction also. “When they do these surveys, they’re, they’re not even...they’re not even going to be thinking about the instructor. You know what I’m saying? They’re going to be thinking about themselves, and that’s ultimately what they’re doing anyways...is when they’re filling out the instructor they’re like how much did I really learn? You know...how, how good am I doing?” These comments echo Dagmar’s concerning the rating process and how it involves as much self-evaluation of how much one has learned as it is an evaluation of the instructor.

Significantly, Sara uses the word “empowerment” when describing what happens when an instructor gives them the sense that they have permission to be honest. She described it as “the process of the student empowerment, what you’re saying in terms of the individual, by

just giving permission...to be honest....” As a consequence, she contended, “...those surveys are going to be honest. At least they’re going to be given the permission to do so.” Whether students take advantage of the invitation will depend on the students and their personal experiences in the class; however, it appears that when the instructor empowers students, they will respond candidly and sincerely.

That students may or may not be candid and/or sincere in their ratings of instructors is clear, based on the results of all the interviews with the participants. Even if students, in general, are predisposed to be honest because of their own personal ethics (as seems to be the case with Dagmar, Sara, Miles, and Asia) or to be cautious about how forthright they can or should be because of their beliefs or uncertainty about how the process operates as well as what the consequences will be (as seems the case with Andrew and Greg), certain factors appear to determine whether students develop a fear of being candid and sincere or develop a sense of trust. Although the administration does appear to play a part in the development of fear or trust for some students, whether students choose to be candid and sincere depends, for the most part, on the instructor. As the discussion to this point has illustrated, depending on various factors, including the comments of the instructor in regard to the rating process itself, the way in which instructors communicate with students, the instructors’ attitudes and personality traits, the manner in which they treat students, and the instructors’ pedagogical skills and practices, instructors have the ability to establish a classroom atmosphere—a climate of fear or a climate of trust—that ultimately determines student candor and sincerity. The core category that emerges from the data is power—the instructor’s power to create an environment in class that promotes fear or trust.

The Core Category: Abuse and Use of Power

Probably neither instructors nor students consciously view their relationship in terms of power. None of my participants ever specifically referred to power. The one participant who even referred to empowerment was Sara, and she tried to downplay that idea. “It has to be, you know...um...you know...an, an, it’s not so much an empowerment thing, you know, because then this sounds radical to me. It just has to be education.” For her, the goal is promoting critical thinking and learning, not empowerment. In commenting about critical thinking and education, she added, “It spills over, and, you know...um...and, and, yes, that is empowering, but I don’t think that’s the goal. That’s what I’m saying. I don’t think it’s about student empowerment because that, it makes it the goal.” Even if empowerment is not the goal from Sara’s perspective, what has emerged from this study is that power and empowerment underlie the dynamic relationship between instructor and students. Depending on how the instructor uses power, the instructor will create a climate of fear or a climate of trust in the classroom.

Whether instructors are conscious of it or not, they possess tremendous power and authority in the classroom. This power that seems inherently afforded to instructors stems from various sources. One source is tradition. From the time children enter school, they fall under the power of the teacher who controls what happens in the classroom. Although later in school some students rebel against the authority of the teacher and of the school, it is probably safe to conclude that the majority of students who make their way to college are students who have acquiesced to the instructor’s being in charge. A second source of power is an evaluation system that rewards or punishes students with grades. Like it or not, we are

still in a society that places a premium on obtaining “good grades.” As Dagmar put it, “They [students] don’t learn to learn. They learn how to make grades.” Grades are the currency to purchase entrance into the elite colleges as well as graduate schools, and the instructor controls the treasury. A third source of power is knowledge. Because of the instructor’s education and experience, most students enter a classroom assuming that the instructor is an authority on the subject matter, a “demigod” as Annie would say, which makes them fearful about saying anything in class that might challenge that authority. Other sources which have instilled the notion that instructors hold the power also exist, but they vary more among individuals. These sources relate to students’ personal value systems that stem from what their parents, their culture, their religion, or other external sources have taught them about respecting the authority of teachers. The point of this study is not to identify the sources and nature of power. Rather, listing some of the sources is merely to acknowledge that most students enter a power structure that is already inherent in the classroom and the educational system. (This discussion is also ignoring the power structure that can appear on totally different levels such as those in connection with patriarchal systems, economic systems, ethnic systems, political systems, and so forth.) For the most part this power structure related to the authority of the instructor is a latent part of the culture of the educational system, but it manifests itself whenever instructors do something to assert their power and authority in class.

According to Schein’s hierarchy of the levels of culture, assumptions, which he identifies as his third and deepest level of the hierarchy, “are the essence of culture” (Baker, 1992, p. 9). Power and control are key assumptions based on learned responses that have already developed over years of conditioning throughout the education process. Because the

predominant manifestation of the power and control dynamic involves a teacher-centered, teacher-in-control structure, this is most likely the environment that students expect upon entering a class. In other words, students likely take a teacher-centered environment for granted. It, therefore, falls upon the instructor to present students with a different environment that shifts from teacher-centered to student-centered and from teacher-in-control to student-in-control. Such a shift will likely transform any assumptions that students may consciously or unconsciously have that they are entering a climate of fear into the reality of an environment conducive to fostering a climate of trust. However, the nature of the relationship between power and climate is not as simple as it sounds.

The core category that has emerged from the data concerns power. The results of this study reveal the complexity of power. First, complexity exists in possible student responses to power. In some cases, students may provide more neutral ratings and no comments for ineffective instructors when they fear the power of the instructor whereas in other cases they overrate ineffective instructors when they fear the power of the administration. Under other conditions, when instructors empower students by demonstrating that their ideas and opinions matter and that they are willing to listen to students, students are more likely—or at least claim that they are more likely—to provide candid and sincere comments as part of the rating process. Additionally, complexity exists in the causes of these responses. As the results of this study show, being ineffective in terms of pedagogical skills and practices may or may not affect how a student rates an instructor. Other factors, such as what an instructor says to students about ratings (or what an instructor does in response to ratings), how an instructor communicates with students, what attitudes an instructor displays in class, and generally how an instructor treats students, all can contribute to the overall climate of the classroom—in

either direction.

Rollo May (1972) claims that, “In psychology, power means the ability to affect, to influence, and to change other persons” (p. 100). By substituting the word “teaching” for the word “power,” one could have an acceptable definition for the act of teaching as well. “Power” and “teaching” are synonymous in the sense that both are transformative acts. May adds, “Each person exists in an interpersonal web, analogous to magnetic fields of force; and each one propels, repels, connects, identifies with others. Thus such considerations as status, authority, and prestige are central to the problem of power” (p. 100). The classroom constitutes such an interpersonal web, in which a power relationship exists between the instructor and the students. The findings of my study demonstrate that the ways in which instructors choose to exercise their power will determine whether the classroom climate is one of fear or trust. However, power relationships also exist at all levels throughout an institution, and how power manifests itself within the institution contributes to the culture of the institution and whether that culture is one of fear or trust as well. Dagmar referred to these levels in her interview in relation to fear: students fearing instructors, instructors fearing deans, deans fearing vice presidents, vice presidents fearing presidents, and even presidents fearing Boards of Trustees. This fear is based on the perception that someone has power and will be exerting that power to the detriment of the person in the subordinate position.

My findings reveal that students can view the power of the administration as a potentially malevolent force either because they know that the purpose of ratings is for instructor tenure or rehiring or because they lack an understanding of what the real purpose is. If this perception occurs on a widespread basis among students, it is probable that a

culture of fear will occur at the macro level, pervading the institution itself. The administration has an important role to play in the process of creating an environment in which students can trust enough to be candid and sincere when completing SRTs. Ironically, even if a climate of trust exists in the classroom, students are still likely to withhold comments or overrate instructors if the culture of the institution is one of fear. The results of this study reveal the impact that student perceptions of administrative response or lack of response can have on student candor and sincerity. However, most students have little or no direct contact with administrators. Consequently, student reactions to them can range from considering them as a potentially harmful force that acts behind the scenes, as seen through Andrew's responses, to seeing them as not only being uninvolved but also irrelevant to what occurs in the classroom, as Miles contended. Because of their immediate contact with the students, the instructor, by far, holds the position of greatest influence over them. Therefore, I will concentrate my discussion of power in the context of the instructor but with the understanding that the same basic principles of power relate to the administration as well.

According to Coll (1986), power manifests itself in three general forms: force, authority, and influence. These forms are not necessarily mutually exclusive. Physical, psychological, or social force involves the imposition of one will over another, often involving constraints and sanctions. Closely related to force is authority, which Coll says involves relationships of power and can take any of the following forms: traditional authority (rule or power based upon established belief in traditions); legal-rational authority (power based on the legality and logic of rules and the legitimacy of the rulers claiming that authority); or charismatic authority (power based on one's ability to inspire devotion through personal magnetism). Finally, influence involves the capacity to produce change insensibly

or invisibly without using force or exercising formal authority. Within both the institution as a whole (the macrocosm) and the individual classroom (the microcosm), the possibility exists for any and/or all of these forms of power to appear.

If instructors are unconscious of their power, they may not even be aware that any of their actions are creating a climate of fear or even that they can choose to create a different climate. Exercising choices is one form of power in and of itself. If instructors are conscious of their power, they can recognize choice and act in such ways as to induce a climate of fear or of trust. Either they can use their power in ways to assert their authority to exert control over the students, or they can share their power by empowering students in various ways to participate with the instructor in the learning process. This act of empowerment involves giving students the power of choice.

I do not want to imply that this dichotomy is simplistic and that asserting authority and control is necessarily bad or unwarranted nor that sharing power will necessarily produce greater learning. For example, if a class discussion is straying on a tangent and the instructor asserts authority by redirecting the discussion, this use of authority has positive ends in terms of keeping student learning directed and focused. In some cases, it may not even be the appearance of authority and control itself that creates the climate but rather the manner in which the instructor deals with students in relation to that authority. For instance, an instructor who, in front of the class, chastises in a humiliating or demeaning manner students who enter the classroom after the class has started is likely to create a climate of fear as opposed to the instructor who deals with the problem privately or with humor (provided it is not sarcastic).

Abdicating authority and control can produce its share of negative results, for example, when an instructor turns class discussion over entirely to the whims of the class. The result could be an entire class period that meanders from one topic to another with apparently no purpose and no clear course or direction, leaving some students wondering what the point was even if they thought it had been a “great discussion.” Although the climate produced is likely one that is relaxed, some students may wonder about the instructor’s ability to teach them anything if all they do in class is have student-generated discussions every time. Furthermore, students who look for a sequential type of class structure are likely to find such an instructor disorganized. Yet instructors who create active and collaborative learning assignments, who engage students in critical thinking activities, who demonstrate a respect for what students contribute, and so forth also foster a climate of trust. As some of the results of this study illustrate, liking an instructor, which probably comes within the context of a climate of trust, can result in overrating in the context of a culture of fear.

Ironically, instructors may induce student learning in a climate of fear just as a climate of trust may end in little student learning. Susan described an instructor she had had this way:

He had a very...stern, rigid...approach to things...um...and could sometimes make comments that could be maybe a little bit sarcastic, and he could come across as maybe overly critical if you didn’t know where he was getting at. But at the same time, I, I felt like...that was preparing me...for...tougher situations later, and so in a sense he was one of the better instructors I’ve had because he was challenging. But as a person, I probably would never, like, want to ha-, go out and have coffee with him or something because of the way his personality...was. He had a way of making kind of almost a belittling personal...personality...maybe I’m much better than everyone else. So as a person I probably wouldn’t like him, and that is probably something he could improve upon as an instructor.

Susan’s remarks also demonstrate that just because students may not like the instructor’s

personality does not mean that they do not recognize what they are learning. So it is not as simple as saying that being sarcastic will cause students to underrate or withhold their comments about an instructor who helps them learn. It is the instructor like Margaret's, whose sarcasm appears in the context of the student ratings the instructor had received from a prior class, who generates a climate of fear that results in students withholding their comments just as Margaret did.

Moreover, these examples point to the paradox that learning or non-learning can occur in either a climate of fear or a climate of trust. Therefore, it is not as easy as providing a formula that would tell instructors that if they follow such and such a course of action, they will produce a climate of fear as opposed to this other course of action that will create a climate of trust, or that a climate of fear will cause students not to learn whereas a climate of trust will cause students to learn. The point is that a climate/culture of fear is likely to produce a lack of candor and/or sincerity whereas a climate/culture of trust is likely to produce the opposite results, and inextricably bound with which climate an instructor creates is how that person chooses to exercise power.

As already mentioned, by its very nature, the educational system is inherently an authority-based system based on centuries of tradition. Besides having traditional authority before they ever enter a classroom, some instructors also exert their legal-rational authority or display charismatic authority as well. What appears to happen in some classes is that during the course of the semester, some instructors resort to psychological or social force when they exert their authority to control students, or they can imply the use of force through behaviors such as cutting off students' comments or ignoring their questions—behaviors that signal to

students that their opinions do not matter and that the instructor controls the power in the classroom. Implied in exerting force/authority to control behavior is the desire to induce conformity. Conformity, in and of itself, is not necessarily bad. It is important, for example, to have students conform to meeting deadlines for submitting assignments or to using standards of writing their papers (or dissertations) such as the APA format. However, some instructors also use force/authority to exact conformity in extreme ways, such as locking the classroom door at five minutes after the start of class to prevent late arrivals, banning the chewing of gum, or other such policies that college students find extreme.

Yet according to Coll, the role of education is to empower people so that they take responsibility: for the direction of their own lives, for empowering or withholding power from their own leaders, and for empowering them to assume leadership roles. She contends that, at times, to achieve this end may mean teaching rebellion rather than conformity, but some instructors may not teach such rebellion, especially if that rebellion may be against the instructor him- or herself. As Coll puts it, “What we [instructors] fear losing is our own power” (p. 422).

This fear of losing power equates with what hooks (1994) describes as “losing control.” According to hooks, this fear of losing control causes instructors to “fall into a conventional teaching pattern wherein power is used destructively.” Such a use of power is their means “of ensuring that the teacher will have absolute authority. Unfortunately, this fear of losing control shapes and informs the professorial pedagogical process to the extent that it acts a barrier preventing any constructive grappling with issues of class” (p. 188). From hooks’s perspective, that fear translates into failing to deal with issues of class. I contend that instructors who exert force and authority to achieve control out of a fear of losing their power

transfer fear, in turn, to the students, with that fear manifesting itself in lack of candor and sincerity in completing SRT's.

May (1972) comments on the power professors have—the power of prestige and status as well as the subtle coercion of others that results from prestige and status. “This [power] is not due to the professors’ conscious aims; it has more to do with the organization of the university and the teachers’ unconscious motivations for being part of it. The more powerless the teacher feels himself to be, the more destructive, even though subtle and covert, will be his influence” (p. 103). May’s observation seems to be a corollary to Coll’s. Instructors who fear losing power may also be, ironically, the ones who feel powerless, thus causing them to exert their power in destructive ways.

However, in the case of students, their fear is not over loss of power but fear of power being used against them. After all, they are the ones in the subordinate position just by the nature of the organization and traditional lines of authority. A profound irony of situation arises when instructors assert their power out of a sense of fear of losing control only to produce fear among those students who are the victims of that power. In other words, fear creates fear. Under such conditions, instructors create a climate of fear, a fear that is rooted in their own fear.

Wheatley and Kellner-Rogers (1996) offer some insight into why people turn to fear as a means of control. They contend that people believe that fear is “the primary motivator for change, that people change only when they are scared” (p. 93). They continue by describing how fear manifests itself:

We have backed ourselves into some fear-filled corners. We have come to believe that to survive, we must control everything. The world cannot be left to itself even for a moment. Fearing people, we control one another mercilessly. Fearing

change, we choose our little plans over the surprise of emergence. We struggle to contain the elemental energies of life, especially people's natural desires to affiliate, to create, to contribute....This world that we seek to control so carefully is a world we have created. We created it by what we chose to notice, by the images we used to describe what we were seeing. (pp. 94, 95)

In effect, these authors seem to be saying that fear becomes a tool for some to use in an attempt to control behavior and to create order. Instructors who wield their power through force and authority are, in effect, disallowing freedom and autonomy in their attempts to promote conformity and control, which may or may not be in the best interests of the students. My findings support the idea that student candor and sincerity depend on the classroom climate that they are in, and the nature of power that the instructor chooses to employ helps determine the nature of that climate.

“It's so easy to walk into a classroom setting and, you know, assume a subordinate role because, you know, the instructor has, you know, Ph.D. or, you know, credentials, and you don't.” Sara's thoughts reflect what many students feel when they enter a college classroom. As I discussed above, the traditional authority of the educational system has already conditioned students to accept their “subordinate role.” Furthermore, this sense of power in relation to superior and subordinate extends beyond the classroom. The entire educational system is a hierarchical power structure, with a superior/subordinate positioning manifesting itself from the Board of Trustees to the administration and from the administration to the faculty, not just from the faculty to the students. In such an environment based on control, students need to “have permission” to provide critical comments long before the ratings occur if the instructor and/or the administration hope to elicit candor and sincerity from the students. Sara's comment is a telling one. Just the word “permission” implies the power relationship that students feel. For instructors to give their permission by making it safe for

students to be critical is an act of empowerment.

Some instructors actively seek to promote this empowerment. They take care in how they exert their authority and tend to exert their power in the form of influence by empowering students through active learning and open discussions rather than trying to control students by shutting down discussion when students are expressing views that they do not agree with as Annie related. As most instructors that try to promote active learning or even class discussion in which students must assume more responsibility have learned, many students do not want that type of power, a topic I discuss in more detail below. Ironically, these instructors must still exert some power to ensure active participation; however, it is the power to influence students as opposed to controlling them. In describing her first experiences in the classroom, hooks (1994) found herself first fearing that she would abuse power, asserting her authority in ways that would perpetuate class elitism and other types of domination. Yet she came to realize that “power was not itself negative” (p. 187). She had to exert some power to “reinforce and maintain coercive hierarchies” (p. 188). As Wheatley and Kellner-Rogers (1996) point out, “We restrict freedom to assert control. We choose control over effectiveness. But living systems cannot be effective if they cannot exercise their autonomy” (p. 102). In effect, instructors should take this advice and be willing to do the inverse: restrict control to assert freedom. Once students begin to assume more power and become more active participants in the learning process, they begin to assert themselves more. They are empowered. Consequently, the climate that the instructor creates in the classroom over the course of the semester is one of trust.

When instructors share their power with students by empowering them through giving them a voice, instructors are more likely to create a climate of trust, leading to candor and

sincerity when students complete SRTs. Emerging from the participants' comments is the finding that when instructors open a dialogue with students and when they act on students' suggestions, especially when the suggestions involve improving the class, students feel valued or, as Sara put it, "taken seriously...honored or respected." Additionally, when instructors display a caring attitude and a genuine concern for student learning and when they practice pedagogical skills such as promoting critical thinking and teaching to different learning styles that lead to more active and collaborative learning in the classroom, instructors are empowering students to be co-participants in creating a climate of trust. Through the promotion of critical thinking and emphasizing learning, the emphasis could well shift from being grade oriented to learning oriented. In turn, the mindset of students might eventually change provided that students are in a true learning-centered institution that promotes a culture of trust.

Sara raised the questions of "what's my incentive for doing this [writing comments]... where's the payoff?" She answered her own questions by saying, "And there's not that it has to be some material reward or, you know, something, you know, umm, something like a grade or something to...something superficial like that, but, I mean, but there has to be some integrity." By empowering students, instructors can provide them with a sense of integrity when completing SRTs, and the students will likely assume their responsibility in the process by providing candid and sincere remarks. The remarks that the participants in this study made generally indicate that, if anything, students will try to find positive ways to indicate areas for improvement and will combine those "negative" comments with others that praise the instructor. Margaret, for example, did indicate that SRTs are a way of thanking an instructor for a job well done. "And the other thing is so often, especially if an instructor is

liked by their students, they can feel that they're doing a great job. And...if, if there are things that the students need to say, especially if they like the instructor, they may feel more comfortable in a form where they get to sit and think about what can I say to help this person. And I think also it can be put to the students that way.” Generally, the results of this study indicate that students want to help instructors improve and to be part of the process as long as they understand the process and feel as if someone will pay attention to their comments and ratings and take appropriate action to correct genuine concerns.

If Coll is right about instructors fearing losing their power, then what Wheatley and Kellner-Rogers (1996) contend about people asserting control while restricting freedom would explain what likely occurs within some instructors and helps produce a climate of fear. When an instructor fears losing power, the inclination is to assert control to create order and restrict freedom. Wheatley (1992) presents what she asserts is a critical paradox that illustrates the relationship that exists between order and freedom:

The two forces that we have always placed in opposition to one another—freedom and order—turn out to be partners in generating viable, well-ordered, autonomous systems. If we allow autonomy at the local level, letting individuals or units be directed in their decisions by guideposts for organizational self-reference, we can achieve coherence and continuity. Self-organization succeeds when the system supports the independent activity of its members by giving them, quite literally, a strong frame of reference. When it does this, the global system achieves even greater levels of autonomy and integrity. (p. 95)

Wheatley is discussing individuals working within organizations (the macrocosm), but these principles seem to apply to students working within the classroom (the microcosm) as well. Instructors who seek to establish order through control and restricting freedom, i.e., limiting student autonomy, likely succeed mostly in engendering resentment and fear in students. This fear manifests itself, in one form, through a lack of candor and sincerity when

completing SRTs. Wheatley's paradox provides the blueprint for what instructors need to achieve—a balance between order and freedom. The order comes not from using authority and/or force to assert control; rather, order emerges from using influence through providing students with “a strong frame of reference” and by empowering them to assert their freedom by expressing themselves, by allowing them to participate in a dialogue, and by teaching them to think critically and to rebel when it is appropriate rather than to conform. In turn, such empowerment is more likely to produce candor and sincerity among all students when they complete their ratings of courses and instructors.

May (1972) describes five types of power: exploitative, manipulative, competitive, nutrient, and integrative. All five can be present in the same person at different times; however, from my experience, the forms that seem to appear most often in the classroom are manipulative, competitive, and nutrient. Exploitative power involves force and presupposes violence or the threat of violence, physical or psychological. Manipulative power is the power over someone. Behaviorists seek to manipulate others, and instructors who assert their power to control student behavior exemplify manipulative power. Competitive power has both a negative and a positive form. In its negative manifestation, competitive power means someone gains while someone loses. May uses the example of student rivalry due to the grading system to illustrate negative competitive power. Positive competitive power appears in team situations involving mutual caring and cooperation in which competition spurs people on to achieve greater accomplishments. Instructors can promote either type of competitive power. Nutrient power is power for the other. May contends, “At its best, teaching is an example” (p. 109). When instructors exercise nutrient power, they seek to help students, to guide them, and to foster growth. Integrative power, May's fifth type, involves

power with the other person. May uses Dr. Martin Luther King, Jr. and Gandhi as examples of how integrative power can transform others through actions that force them to examine their own consciences. Another example, however, involves someone who seeks the criticism of others to improve him- or herself. When instructors reach a point where they seek criticism because they know that they can use it to improve, they are exerting integrative power. Coll's categories of force and authority seem to encompass May's exploitative, manipulative, and competitive power categories; influence appears to align with nutrient power and integrative power.

Although not all of my participants cited specific examples of instructors, the examples that they gave demonstrate all the forms of power. No one gave an example of an instructor who used physical force; however, Miles's example of the instructor who ignored student questions and created an atmosphere that caused students to withdraw probably comes the closest to exploitative power. Dagmar's example of the instructor who awards points to students who complete an evaluation illustrates manipulative power. From her perspective, the instructor was someone who was trying to "buy" positive evaluations by giving extra credit points. Also, Annie's example of the instructor who cuts off student responses when she does not agree with them displays this form of power. Perhaps Margaret's example of the instructor who uses sarcasm about prior student ratings illustrates competitive power at work through the act of belittling the comments of prior students. In this case, the instructor behaves as if she is in competition with the students. The description that John provided of the instructor who used various teaching techniques to appeal to different learning styles probably best exemplifies the use of nutrient power. Integrative power seems to appear in Andrew's example of the instructor who sought student comments on how to improve the

class and then acted on those comments.

Nutrient power is likely more easily achieved for most instructors and integrative power is possibly the most difficult because integrative power involves exposing one's ego to criticism. May provides an excellent description of what can happen, and although he is not specifically describing a process directly related to teachers, it is easy to see how the description applies:

Our narcissism is forever crying out against the wounds of those who would criticize us or point out our weak spots. We forget that the critic can be doing us a considerable favor. Certainly criticisms are often painful, and one has to brace one's self in the face of them. We can slide back into manipulative power (by forcefully silencing the critic) or competitive power (by making the critic look silly). Or we can even protect our thin skins by means of nutrient power (patronizing the critic by implying he is confused and needs our care). But if we do regress in these ways, we are losing an opportunity for new truth that the questioner, hostile or friendly as the case may be, may well be giving us. (p. 110)

The instructor who uses integrative power, particularly in the form of seeking criticism, is rare.

Coll contends, however, that exercising nutrient or integrative power does not ensure empowering others necessarily. "In the very process of dependency, people learn to abdicate responsibility to others and therefore escape responsibility for the consequences of their actions. There is security in being taken care of and it is easier to have someone else responsible for consequences" (p. 418). Coll refers to Paulo Freire, who labels this a "fear of freedom." Students have been involved in a system of authority that has sought to control them. It may be not only easier but also safer for some students to avoid being empowered. While many students may be content having instructors provide them with information that they dutifully write in their notebooks, memorize, and regurgitate on their exams, some students may desire that type of environment because they want to keep a low profile. I know

from firsthand experience that this occurs. Recently, an older female student came to my office in tears because of how a female instructor was treating her on a daily basis in class, mispronouncing her name and belittling her comments to the point where the student said that she sits in a corner seat at the back of the room, keeps her eyes on her notebook, and spends each class period praying that this instructor will not call on her. This example illustrates that although some students may simply want the instructor just to spoon feed them the course content, others avoid being involved out of fear.

Additionally, Coll cites the experiments of Stanley Milgrim, in which subjects willingly tortured an innocent person with what, they believed, were increasingly stronger electric shocks because an authority figure urged them to do so. “Among the forces which lay the groundwork for obedience, Milgrim lists family, impersonal systems of authority and repeated rewards for compliance or punishments for failure to comply” (p. 420). Students may be more likely to respond the way they do in general because of the conditioning they have received over time.

Ironically, just as students can be victims of the authority of the educational system, so, too, can instructors. As Coll points out, “Perhaps no other profession except parenting receives so little preparation. Educators act as if leadership is a by-product of education. But experience shows that when persons assume authority they often replicate their own experience of authority just as parents replicate their own experience of their parents. They have no theory or norms by which to guide themselves” (p. 422). In other words, instructors tend to perpetuate not only the pedagogy under which they probably learned best but also the behavior that they may associate with the role of instructor, which may be one that supports the use of force and authority to maintain strict control.

Add to that irony the paradox that in any given classroom some students may fear control and repression while others may fear freedom. As a result, instructors must engage in a dual process of being self-consciously working on changing themselves and the type of learning climate that they create in the classroom while trying to transform students into active learners. Because a classroom is likely to contain both students who want to be engaged along with students who prefer to remain disengaged, instructors who seek to empower students must balance exercising nutrient and/or integrative power with using a degree of manipulative power to engage students who would prefer not being empowered. The key to success rests in how instructors approach the problem. Those who use authority and force are more likely to generate fear while those who try to manipulate through using influence are more likely to generate trust. It appears to me as if instructors who exercise integrative power and open themselves up to criticism and then act on that criticism would have the greatest chance of success because the consequence of their actions will likely be to create the climate of trust in which students will feel safe in becoming involved and developing the types of responsibilities that result from empowerment that Coll listed. A climate of trust is one that will probably promote change in the students and lead to candor and sincerity in their ratings.

From the time they enter the class on the first day, instructors have choices about how and when they will exercise their power. Some instructors choose to distance themselves from students and have students address them as “Dr. Doe.” They believe that a degree of formality is critical for guaranteeing respect and ensuring appropriate behavior on the part of students. Other instructors choose a more relaxed style and have students address them as “Jane” or “John.” Yet power exists in more than the formality or informality of titles.

Instructors can also control the power relationship by asserting their authority through force and control or by sharing their power by empowering students. According to Bolman and Deal (1995), “Control is an illusion. It’s seductive because it gives a feeling of power. Something to hold on to. So it becomes addictive. It’s hard to give up even when it’s not working” (p. 31). The paradox of this situation is that by influencing students to assume some of the control through empowering them, instructors will actually obtain the control that they were so afraid of losing when they use force or authority to exert control. As Bolman and Deal add later, “You can give power away and wind up with more” (p. 92).

It strikes me that when anyone within an organization uses power in the form of force and/or authority to establish and maintain control and to dominate others, this behavior puts distance between people with some of the consequences being the creation of fear and mistrust. “Organization” can include both the institution as a whole or the classroom. As I have already mentioned, in such an environment in the classroom in which instructors seek to control students, two fears can emerge—the fear of losing power by the instructors and the fear of being punished in some form by the students. Fear, then, can create a self-perpetuating cycle that creates a climate of fear and mistrust. Wheatley and Kellner-Rogers (1996) contend that, “Organizations can keep searching for new ties that bind us to them—new incentives, rewards, punishments. But organizations could accomplish so much more if they relied on the passion evoked when we connect to others, purpose to purpose. So many of us want to be more. So many of us hunger to discover who we might become together” (p. 63). Perhaps if instructors could use their power to influence rather than control and if they could engage in the integrative power that May speaks of and open themselves up to constructive criticism from students, they could discover as much about themselves as

students learn. If instructors also apply their influence to empower students rather than merely control them, the entire classroom environment will transform from one of fear to one of trust. It is ultimately this climate of trust that will likely enhance the chances that students will be candid and sincere when they rate instructors and classes.

Ultimately, abuse and use of power in the classroom and, by extension, in the institution deserve more study. Just as fear and trust are the two extremes of a continuum, so, too, are force/authority and empowerment two extremes of another continuum. Just as one cannot classify students as either fearful or trusting, one cannot oversimplify the role of the instructor as either an authoritarian whose power grips students in a stranglehold of control or a facilitator who benevolently nudges students through empowerment. Rather, it appears as if a relationship exists between the forms in which power manifests itself and the extent of fear or trust that an instructor generates in a classroom. Depending on how students perceive that power will determine, to a great extent, whether they find themselves in a climate of fear or a climate of trust. The more fear they feel, the more likely they are not to be candid and/or sincere in completing student ratings of teaching. The more trust they feel, the more likely that the opposite result will occur. This co-relationship between the uses of power and the climates that result is just one area that deserves further research.

CHAPTER FIVE

DISCUSSION

Despite a myriad of studies related to the reliability, validity, and generalizability of student ratings of teaching, two findings troubled me throughout review of the literature: although the preponderance of studies report that for the most part SRTs are reliable, valid, and generalizable, a minority of studies still present evidence that biases and factors such as the halo effect exist, which, in turn, casts doubt upon the reliability and validity of SRTs. One cannot be certain whether a factor such as the likeability of the instructor, for example, results because of a true bias on the part of the rater or because it is just one of the byproducts of being an effective instructor. Furthermore, several researchers (Abrami, 1985; Abrami, 1989a; Baker & Copp, 1997; Cohen, 1983; Feldman, 1997; Kulik & McKeachie, 1975; McKeachie, 1997) caution institutions about using the results of student ratings for decisions concerning tenure and promotion. Consequently, despite findings that SRTs are generally reliable and valid in relation to the instrument, those examining the results should take care in how to interpret them, particularly in their use for tenure and promotion, thus casting doubt on overall reliability and validity. In other words, the message from these and other studies might be summarized by the following paradox: SRTs are reliable and valid to the extent that the instruments have been deemed so but not reliable and valid enough to guarantee trustworthiness.

The second finding that troubled me was that almost all studies direct their attention to

the rating instrument itself—to the uses of the rating instrument (consequential validity), to the issue of dimensionality (conceptual structure), to the correlation of the instrument with indicators of effective teaching (convergent validity), or to factors that may indicate biases that will impact the reliability and validity of student ratings (discriminant validity). This attention to the instrument is understandable since controlling the instrument itself, which is static, is far easier than controlling human beings, who are dynamic. However, studies that aim at probing what students are actually thinking when they complete student ratings—of simply asking students whether they are being candid and sincere when they complete a rating of an instructor and course—apparently do not exist.

A study by Kishor (1995) represents one of the first to begin moving toward analyzing cognitive processes involved in producing numerical ratings. As Kishor points out, if the operation of implicit personality theories that affect the cognitive processing of performance information results in biases in performance ratings, then even psychometrically sound rating instruments will likely fail to provide valid evaluations. In other words, even a valid instrument might fail in providing reliable results because of invalid student rating judgments. The paradox of this conclusion is almost overwhelming in its implications.

Robert Creeley's poem "Mind's Heart" expresses this paradox nicely:

Mind's heart, it must
be that some
truth lies locked
in you.

Or else, lies, all
lies, and no man
true enough to know
the difference.

The question then becomes, "If some doubt exists about the reliability and validity of the

results of SRTs even from valid instruments, should the process of asking students to rate instructors and classes continue or not?” My answer is, “Maybe...it depends on how power manifests itself and the resulting climate in the classroom and perhaps even the culture of the institution that the assertion of power creates.”

Based on the results of my study, it appears that student candor and sincerity depends on whether they are responding primarily from within a climate of fear or of trust at the micro or classroom level as well as possibly from within a culture of fear or of trust at the macro or institutional level. Which climate or culture emerges hinges on whether the instructor or even the institution asserts power in ways that students perceive as threatening in some manner or whether the instructor shares power by empowering students. Yet I need to stress that even this dichotomy is a generality. Likely within both climates/cultures of fear or trust are students who will be candid and sincere and others who will not be for various subjective rather than objective reasons. In fact, student candor and sincerity is more likely on a continuum, with being completely candid and sincere at one extreme and totally withholding comments and being insincere at the other extreme. Most students likely fall somewhere between, providing varying degrees of candor and sincerity, depending on the degree of fear or trust that they feel with these two factors being poles of a continuum. Consequently, no means may exist of truly knowing whether one can ever trust the results of SRTs when doubt exists about student integrity and that, at best, SRTs probably provide a general sense of instructor effectiveness. Furthermore, a culture of fear at the macro level may even override a climate of trust at the micro level. Although it is beyond the scope of this study, the culture of the institution (the macro level) likely impacts the amount of fear or trust faculty members feel in relation to the process, which, in turn, can impact the climate these instructors create

at the micro level in the classroom. However, the core category related to the abuse and use of power that has emerged from this study indicates that how an instructor applies power within the classroom can influence the climate, and within that climate, students will tend to fear or trust being candid; therefore, the likelihood of trustworthiness decreases or increases in direct relationship to the degree of fear and trust students feel. The fact that these two predominant yet diametrically opposed climates appear to exist may explain, at least in part, the inability of research to prove conclusively that student ratings are reliable and valid.

The findings of this study both negate and support some of the studies presented in the review of the literature in chapter 2; however, this study does not resolve the issue about reliability and validity of SRTs. Instead, the findings demonstrate that the participants admitted to both sets of phenomena: a lack of candor and/or sincerity under certain circumstances and total candor and/or sincerity under others. By probing for the reasons causing these phenomena, I discovered that I could trace them to an underlying feeling of fear or a feeling of trust. However, the findings also revealed that through building trust, these students are more likely to produce the second set of phenomena by rating and commenting on instructors with candor and sincerity. Underlying both fear and trust is the core category of power, for it is through the exercise of that power that instructors can create a classroom climate and institutions can create a culture of either fear or trust.

Since much of the literature reviewed in chapter 2, particularly the sections on the conceptual structure and on convergent validity, relates to proving the reliability, validity, and generalizability of the rating instrument itself, many of the studies presented in chapter 2 fall beyond the scope of the findings in this study. To a very limited extent, some of my findings negate a few studies or both negate and support parts of other studies in relation to

consequential and discriminant validity. The findings of this study support the literature in these two areas of validity to a greater extent than they negate it; however, the majority of my findings extend beyond what I have found in the literature, particularly in regard to the core category of power. In the remainder of this chapter, I will summarize and interpret my findings in relation to how they negate and support the literature followed by how my findings extend beyond the literature. The chapter concludes with sections discussing the implications of this study and making recommendations related to the application of this theory for instructors and institutions along with suggestions for further study.

Summary of the Findings That Negate or Partially Negate the Literature

The great paradox that I discovered in the course of interviewing my participants in relation to the phenomena of candor and/or sincerity (or the lack thereof) is that all 12 of them indicated that they are generally candid but that they had either withheld their comments and/or had overrated instructors on occasion. Only Miles and Jaime of the 12 participants never mentioned a lack of candor, but both of these participants admitted to overrating certain instructors. Although none of the participants conceded to intentionally underrating an instructor, John acknowledged that he would underrate an instructor if he knew that other students were intentionally overrating that instructor. Of course, the problem for instructors and administrators, based on the implications of these findings, is that if students are not consistently candid and sincere in completing SRTs, neither instructors nor administrators can trust the degree to which results are trustworthy—that is if they are trustworthy at all. The unspoken assumption that appears to underlie the studies of the instrument is that even if certain biases may affect student ratings, students will be generally candid and sincere in their ratings, but the findings of this study reveal that this is likely a

faulty assumption.

In fact, various conditions could invoke various responses as revealed in my findings. In circumstances in which students fear being identified and/or fear some form of instructor retaliation, they may or may not be sincere in terms of their ratings on Likert-scale items, but they will almost certainly not be candid and/or sincere with end comments if they provide any at all. In circumstances in which students fear some sort of administrative response that may result in punishment for the instructor (disciplinary action/non-renewal/non-tenure) or punishment for the students (class cancellation/denial of credit), students will likely overrate instructors on Likert-scale items and provide either no comments or insincere comments. Finally, in circumstances in which students “fear” that the administration will not respond, thus resulting in a general sense of apathy, students are likely to give instructors neutral ratings on Likert-scale items and provide no comments.

Furthermore, as Miles pointed out, another student response that has nothing to do with ratings can occur. Some students may be rating a class or an instructor with their feet by withdrawing from a course in which they are having problems with the instructor or when they feel that an instructor is not teaching them anything. Unless institutions survey students who withdraw from classes, it is impossible to know whether the reason for the withdrawals is instructor related or personal. Even if patterns showing high student withdrawal rates appear for a particular instructor, an institution cannot be sure whether those rates indicate legitimate problems with the instructor or demonstrate that some courses are more demanding, causing those students who are looking for an “easy” grade to withdraw in larger numbers.

Anecdotally, I can present yet another possibility. Recently, a student came to my

office to report the behavior of an elderly female instructor who constantly belittles students and student work in class the semester after he had completed it. The student reported that many students simply withdrew from the class over the course of the semester. When I asked why he or others had not reported this behavior when it was occurring, he said that he did not want to say anything that would cause this instructor, who is an adjunct faculty member, to lose her job. The student was reporting the behavior now only because he had related both his first-hand experiences as well as those involving other students to another instructor who had then urged the student to report this behavior to the dean. In the case of this student, it appears that it is easier to be a silent victim than it is to feel guilty about causing someone to lose her job. Other students did not report any problems perhaps for similar reasons but just as likely because of either fear of retaliation or fear that no one would do anything about the problem, so why bother? (It is even possible that some students share Andrew's attitude about future students needing to endure what he had had to endure.)

What these responses indicate is that fear likely exists for students even after they withdraw from a class. I imagine that the students who withdraw are simply relieved that they escaped further mistreatment and that the culture of fear is so pervasive that it is easier just to walk away to avoid further involvement. Such instructors as the one I just described are certainly the exception, but the fact that instructors like this do exist and that the student recourse of choice is either to suffer in silence or withdraw without notifying anyone provides another reason for having doubts about how trustworthy ratings are. It may be that through mistreatment of certain students, a few instructors succeed at the micro level in coercing them into withdrawing, leaving a smaller group of students who have no problems with the instructor and give those instructors higher ratings. Then, too, in an environment in which

students emphasize grades rather than learning and in which at the macro level the institution asserts its power by establishing that the primary purpose of SRTs is for tenure and/or rehire rather than to improve teaching and learning, a variety of situations are possible that can undermine the trustworthiness of the results of a generally reliable instrument.

My findings as well as experiences like the anecdotal encounter I reported in chapter 4 support Kishor's contention (1995) that it is possible to have untrustworthy results even though the instrument itself is reliable and valid. Consequently, this general finding throws into question all the literature concerning the overall reliability and validity of any SRT when these considerations are removed from the context of examining the instrument and the results that the instrument produces and examined in the context of whether students are candid and sincere when providing their ratings.

Besides my findings generally casting doubt on the reliability and validity of the results of SRTs despite the reliability and validity of the instrument, the response of one participant appears to refute the findings that Marsh (1987) and Costin, Greenough, and Menges (1971) report showing that, in regard to the impact the reason for taking the course has on the ratings, general interest emerges as the only reason to have a substantial effect on ratings and that required courses receive slightly lower ratings than elective courses. Only one of my participants (Greg) even mentioned distinguishing courses that he was interested in from general education courses he was required to take. His response, however, was not that he necessarily gave general education courses or instructors of those courses lower ratings. He indicated that he generally rates them with more neutral responses. However, he claimed to be a much harsher critic of courses he chose because of his interest in the subject matter. With these courses, he gives low ratings when the course fails to meet his expectations for

learning or highly when they do. Overall, while it probably is generally true that required courses receive slightly lower ratings than elective courses, my findings indicate that student interest in the subject matter is a non-factor as no other participants made any distinction between courses on this basis.

The most important negation of other research comes with the issue of fear of retaliation. As noted in chapter 2, I found that, at most, the literature occasionally mentions the need for having a neutral person administer the ratings and ensuring student anonymity to remove or mitigate any possible student fears. Miller (1988) comments about student suspicions about ratings that an instructor personally distributes and collects in class, fearing that the instructor will recognize the source of any critical comments and may be vindictive towards any critics of his or her teaching. My findings definitely support Miller's contentions, but they also demonstrate the issue of fear is far more extensive than Miller presents. More significantly, my findings contradict Spencer and Schmelkin (2002). They studied student perspectives on teaching and evaluation and found that students show little fear of retaliation. Nine of my 12 participants revealed some degree of fear of retaliation as a major factor in determining whether they would provide candid comments as part of an SRT. One possible reason for this discrepancy may link to my finding that students do not always fear retaliation and that, in general, all my participants indicated that they are usually candid and sincere. It is possible that the students who responded in Spencer and Schmelkin's study, like my participants, are usually candid and may generally not fear retaliation. Additionally, the college from which my participants came conducts faculty ratings around mid-semester, well in advance of the end of the semester. Fear of retaliation may be greater when ratings occur at times other than the end of the semester or during the last class. Another reason

could be that the predominant culture of the college from which I drew my participants is one of fear whereas the students responding to Spencer and Schmelkin may have come from a culture of trust.

Moreover, I see fear as related to a larger “system,” which emphasizes the importance of grades. “System” could be synonymous with “culture.” Fear is more likely to occur in an environment in which people believe they have something to lose. In an environment that uses ratings for the purpose of tenure, promotion, and rehiring, instructors will fear losing their position or their job. In a grade-oriented environment, students fear losing a high grade. Most students perceive grades—not learning—as being the currency to buy admission into four-year schools and into graduate schools regardless of whether this perception is true. It is understandable in such an environment for students to be fearful even if that fear is not warranted. Ironically, while the fear connected with the instructor (retaliation against the student) or with the administration (retaliation against the instructor) differs, the result is the same—students fail to indicate their concerns about an instructor’s inability to optimize student learning. This irony indicates that the issues of reliability and validity, particularly in connection with fear of retaliation, may be far more complex than the literature would lead one to believe. Depending on the conditions that exist, students may be reacting to a fear of the instructor or a fear of the administration but for totally different reasons. How would anyone really know? More study needs to occur in relation to the climate of a classroom and the culture of a college to determine whether such a distinction exists and can account for the discrepancy in these findings. However, their findings and mine agree that those students registering concerns about retaliation are also those who are more reticent about completing ratings.

The results of my study also partially negate and partially confirm the list of factors that Marsh (1982b, 1987) reported as being ones that faculty members believed would cause student bias as well as the factors that Feldman (1988) reported concerning faculty and student views of characteristics of good teaching. Of the factors Marsh listed as receiving the highest percentages from instructors, my findings negate course difficulty, grading leniency, and prior student interest as being factors that my participants consider in terms of ratings and support instructor popularity and reason for taking the course. Additionally, my findings support other research that concludes course workload not to be a factor that creates a negative bias. In terms of Feldman's findings, the results of my study negate the quality of being clear and understandable, which appeared as a high value both for faculty and students, as well as elocutionary skills and teachers being interesting, which students highly valued. On the other hand, my findings support the factors of being prepared and organized, being sensitive to class level and progress, and being open to the opinion of others, all of which both faculty and students value.

None of my participants indicated that course difficulty had any influence on their ratings. John feared that some students might want to give instructors lower ratings because "somebody that never showed up for class and all of a sudden was getting a failing grade because they didn't deal with their attendance and they're going to like beat this up with the instructor because the instructor's following the policy that's there and that every student should know about." However, in this case it is not the difficulty of the course that may be a factor but rather some students' lack of motivation (or the instructor's failure to motivate them) to attend class or to do the work combined with receiving a low grade that may cause them to want to retaliate against the instructor by means of the ratings. The belief that such

an action could occur makes sense in a grade-oriented culture because a student with a low grade has nothing to lose even if identified, therefore not merely being candid but actually intentionally underrating the instructor.

What John sees as a potential student reaction (one that he said that he would counter by overrating the instructor) supports the 1950 study by Elliott, which Kulik and McKeachie (1975) cited, in which Elliott discovered that students respond according to their ability levels in relation to whose ability level the instructor teaches to. In Elliott's study "bright students" rated instructors who taught to their level more highly than instructors who taught to the level of "low-ability" students with the inverse being true with low-ability students. Elliott's findings come closer to explaining the response of the students John is referring to than the contention that students are lowering their ratings because of the difficulty of the course. Asia's desire to have instructors who are willing to take the time "to just break it down even though we are on a time schedule" also appears to point to this factor of teaching to the level of the student. It is easy to imagine in a class in which an instructor moves quickly through complex material because the so-called bright students are able to work at that pace would receive high ratings from those students who can keep pace but lower ratings from students who cannot. One other factor that may account for low student ratings in connection with the grade relates to the findings of Dowell and Neal (1982) in conjunction with attribution theory, according to which poor students will attribute their performance to external factors, such as the instructor. A combination of students being grade oriented and either being more inclined to underrate instructors who do not teach to their level and/or being more inclined to blame the instructor for their poor performance in accordance with attribution theory than because of course difficulty appears to explain better the potential student reactions that John

suspects could occur. Future research may want to revisit Elliott's findings in connection with student attitudes about grades and attribution theory.

The closest that any of my participants came to indicating that students rate easy graders more favorably did not occur in the context of actually receiving a higher grade than deserved and therefore giving the instructor a higher rating. It came in the example that Dagmar gave concerning the instructor who gave bonus points to students who completed a course evaluation that included the students' names. From the incident that she described, the implication was that this instructor was "buying" higher ratings through awarding bonus points for responding. Asia also revealed how she has discussed instructors' reputations with other students concerning whether the instructors are demanding or lenient. One might assume, therefore, that easy graders receive higher ratings; however, my findings do not support this conclusion. Based on Asia's comments, it seems more likely that whether an instructor is an easy grader plays a bigger role in course selection, not in how students rate instructors. None of the participants in my study indicated that they overrate easy graders or even reward instructors for receiving high grades.

As for the factor of prior student interest, only Greg mentioned his interest in the subject as having any bearing on his ratings. However, rather than prior interest serving as a positive biasing factor, his prior interest causes him to be more critical of instructors in these courses. If anything, Greg is more likely to rate instructors lower for failing to meet his expectations rather than reward instructors merely on the basis of his prior interest. None of the other participants even referred to prior interest.

In his meta-analysis of comparing faculty and student opinions about the characteristics of good teachers, Feldman (1988) found that both faculty and students value instructors who

are clear and understandable, both place moderate value on being open to discussion, and students, not instructors, place high value on elocutionary skills and the instructor's being interesting. None of my participants specifically mentioned any of these characteristics. They did, however, mention organization and following the syllabus more often than any other factor that they consider when rating instructors. Being organized is one of the characteristics that both faculty and students value according to Feldman's meta-analysis. It is possible that organization may be synonymous with clarity for some students, which may account for why they did not mention that factor. Sara's emphasis on instructors' promoting critical thinking in the classroom implies valuing class discussion, but Sara did not specifically refer to class discussion as being a factor she considers. Evidently, the method of instruction plays little if any part in student ratings of my participants because none of the participants mentioned that they rate a particular teaching style more highly than another although John did praise the instructor who teaches to multiple learning styles and works to promote active learning. Diane mentioned instructor enthusiasm and ability to excite the students about the subject matter, and it is likely that the instructor's energy level contributes to whether students view the instructor as interesting. However, it is just as likely that instructors' choice of subject matter and the degree to which they actively involve students in the learning process could result in the students' viewing those instructors as interesting. None of the participants specifically mentioned the interest that instructors generate.

My findings do support some of Marsh's (1982b) list of factors that faculty members believe are student biases as well as some of Feldman's (1988) list of characteristics of good teachers according to faculty and students. Instructor popularity was third on Marsh's (1982b) list with 63% of the faculty believing it to be a biasing factor. As reported in chapter

4, Annie, Susan, and Jaime commented on overrating instructors because they liked them. Although instructor popularity and likeability are not exactly the same, one can assume that likeability contributes to instructor popularity. Interestingly, liking an instructor and overrating that instructor occurred in a culture of fear—a fear that the administration will dismiss or punish the liked instructor. As also reported in chapter 4, students are just as likely to exercise candor and sincerity in their comments through suggesting improvements even for instructors they like as John’s comments indicated. Additionally, as the literature points out, being popular can easily result from being a good teacher, in which case instructor popularity is not a biasing factor. Thus, my findings indicate that liking an instructor is likely to be a biasing factor in rating a less-than-effective instructor when students are in a culture of fear but may well not be a factor leading to biased ratings in a culture of trust. Knowing that the purpose of the ratings is for improving teaching and learning as opposed to tenure and job retention could make a difference in causing students to give constructive criticism even to instructors they like but who could improve in certain respects.

Only Diane and Asia mentioned anything related to the workload in courses; however, neither of their remarks indicates that the amount of work in a class necessarily affects the way they rate instructors directly. Diane’s complaint is not the amount of work but rather when instructors assign more work than they are willing to mark and return in a timely manner. As Diane said, “I don’t mind, uh, the amount of work as long as I feel I’m getting feedback from the work getting, returning it.” Asia also did not discuss workload in the context of ratings. She was merely discussing the practice of some instructors, which she described this way:

I mean, it’s all good for us, you know, to work hard and be prepared for the

university that we are trying to go to. Um, but the same time, I don't know if they take into consideration that we are taking other classes, but that's what all adulthood is about. But I just feel like some teachers give *a lot*...not because they...some people ha-, most instructors do it 'cause they have to...meet the requirements, and then some teachers go *overboard*...overload. And I don't think they take into consideration of...um...can a, can a student produce...good work...not just how much work they can get done.

Asia's comment came in the context of discussing the value of SRTs and being able to relate that information to the instructor and/or to the administrators. In her case, she was not indicating that a heavier workload would create a bias in her ratings. She even asserted that having a heavier workload is "good for us" to be prepared for transfer to a university, and in her comments about demanding instructors, she associated more learning with the demanding instructors. If anything, both Diane and Asia's comments support the research that ratings for classes with more demanding workloads tend to be higher or no different (Freedman & Stumpf, 1978; Frey, Leonard, & Beatty, 1975; Marsh, 1980; Marsh 1982a; Marsh 1983).

Reason for taking the course, which 55% of the faculty reported in Marsh's findings believe is a biasing factor, appeared as a consideration for only for Greg. He admitted that he tends to rate instructors in required general education courses more neutrally than the instructors in classes he has chosen to take because of his personal interest. His report that he also tends not to write any comments in these classes, however, indicates that he does not have as much of a negative bias toward general education classes as he has a general sense of apathy about rating them. Greg made it sound as if he really did not care that much one way or the other about how he rated courses the institution requires him to take. However, his remarks clearly demonstrate that he differentiates his reasons for taking the course and that this differentiation affects how he undertakes the rating process, thus indicating a bias. The fact that he is the only participant to mention reason for taking the course as a factor indicates

that, in general, if the reason for taking a course is a biasing factor, it is likely not a prevalent or conscious one among students.

Of Marsh's participants, 53% viewed student's GPA as a potential bias. Apparently, faculty believe that students with higher grade point averages may rate courses higher and those with lower grade point averages may rate courses lower. In this respect, students' GPA's may somehow be associated with the factor of grade leniency. Ten of my 12 participants have GPA's over 3.0 and half of those have GPA's of 3.5 or higher. My findings indicate that GPA has no bearing whatsoever on student ratings. One of my major conclusions is that all of the participants will be either candid and sincere or not depending on the instructor and the climate or culture. Lack of candor and sincerity can occur with any student, regardless of GPA, thus indicating that, at least for these participants, GPA is not a biasing factor.

The only other factor on Marsh's list is class size. Although class size may certainly play a factor in large universities, community colleges pride themselves on trying to avoid overly large classes. None of my participants mentioned class size, and it is doubtful that any of them consider class size when completing their ratings.

My findings confirm other characteristics that faculty and students value in good teachers, according to Feldman's findings. These include characteristics that both groups value, including teachers being prepared and organized, being sensitive to the class level and progress, and valuing the opinions of others. Jamie, Sara, Annie, and Asia all mentioned organization as a factor that they consider. In all four cases, organization became a concern when instructors were not organized or did not follow their syllabi as opposed to the participants citing examples of instructors who were well organized. Apparently, participants

take well-organized classes for granted but particularly notice disorganized ones.

As cited earlier, Asia showed a concern about instructors not showing sensitivity to student progress in relation to how quickly they cover material in class. Additionally, John's example of the instructor who teaches to multiple learning styles presents an instructor who illustrates sensitivity to class level and progress although John did not discuss this instructor in those terms. In both cases, instructor sensitivity to student progress connects with being an effective instructor, at least from the participants' perspectives. More importantly, such treatment of students demonstrates the type of caring attitude that I contend promotes a climate of trust. Whether it also constitutes a bias or is a true indicator of effective teaching is something I cannot determine in the context of this study.

The final characteristic that both faculty and students agree as being an important characteristic of good teachers is valuing the opinion of others. Part of valuing the opinion of others must come from eliciting these opinions and listening to what students have to say. Dagmar, Sara, Annie, Miles, Jamie, and Susan all provided evidence of the importance of communication to them and how it plays a key role in creating a climate of fear or trust in the classroom. When instructors fail to engage in a dialogue with students or when they fail to answer student questions or cut off student answers when they disagree with the student or appear not to like the students' answers, they send the message to students that they do not value or want to hear students' opinions. Students need to feel comfortable talking with their instructors both in and outside of class. My study confirms the key role that valuing the opinion of others, as reflected through the communication process, plays in affecting a climate in the classroom that will aid in promoting candor and sincerity.

Although the results of my study appear to cast some doubt on the trustworthiness of

the results of SRTs even when the instrument itself may be reliable and valid, in general, my findings present results that contradict the literature only in relation to the reason for taking the course as being a biasing factor and to the fear of retaliation being less significant than I found it to be. Additionally, my findings both negate and support the research concerning faculty beliefs about biasing factors and concerning the characteristics of good instructors that faculty and students hold. More of my data support aspects of the literature than negate it.

Summary of the Findings That Support the Literature

At the beginning of this chapter, I noted that my findings relate mostly to the aspects of consequential validity and discriminant validity and not to studies concerning conceptual structure and convergent validity. Although my findings have only a limited connection with the issue of consequential validity, the core category that emerges from my study supports some of the literature related to consequential validity. Finally, the literature presents characteristics associated with superior college teachers and teaching that my findings also support.

Findings Related to Consequential Validity

Kulik and McKeachie (1975) contend that ratings for the purpose of tenure and promotion may impact the general academic community atmosphere (i.e., use of ratings may create an atmosphere of distrust, hostility, resentment, etc.). My study focused on factors affecting student behavior in relation to completing ratings. However, based on the description of some of the instructors with whom my participants have taken classes, these instructors clearly displayed hostility and resentment. In particular, Margaret's instructor who made sarcastic comments about prior student ratings comes to mind. Also, the examples

of Annie, Andrews, Susan, Greg, Jamie, and Diane in relation to their fears of administrative responses support the atmosphere of distrust of which Kulik and McKeachie must be speaking. In essence, Kulik and McKeachie are describing what I am labeling as a culture of fear.

As part of his discussion of consequential validity, McKeachie (1997) proposes the faculty need training in how to instruct their students on becoming more sophisticated raters. He also suggests that faculty explain to their students why ratings are likely to result in better learning. Sara's comments about the need for instructors to teach students to apply critical thinking skills throughout a class would represent one idea to lead to making students more sophisticated raters. As my findings demonstrate, students show a sense of responsibility and maturity when undertaking ratings as seen in the comments of Dagmar, Sara, Susan, and Asia, in particular. They appear both ready and willing to undergo training in being better raters. Additionally, my findings, as demonstrated through comments from Andrew, Asia, and Susan, reveal that when students know that the purpose of ratings is to improve teaching and learning, they indicated that they would be more likely to be candid and sincere when completing SRTs. Although requesting that students be honest in their ratings is no guarantee, in and of itself, of promoting more candor, participant comments indicated that when instructors explain the purpose of the ratings and ask for honesty, they are more likely to promote that response. Of course, it helps when instructors have already established a climate of trust. What will enhance the likelihood of receiving candid and sincere responses is providing students with a sense of how the instructor plans to use the comments and/or has used the comments in the past to improve teaching. Better yet are the instructors, like the instructor Andrew described, who solicit student comments early in the semester and enact

some of the suggestions, thus demonstrating to students that they are taking student comments seriously and are showing their commitment to the process through their actions, not just their words.

My findings support Kulik and McKeachie's (1975) contention that administrators should use caution when employing ratings for the purpose of promotion, tenure, and, by extension, job retention. When the instructors feel as if their jobs are on the line, that situation can engender a level of fear in them that is comparable to the level of fear students feel when their grades are on the line. Unless instructors and institutions convince students that the primary purpose of ratings is to improve teaching and learning, many students will mistrust the process and fear either an impact on their grade or their workload or a negative repercussion for the instructor should they indicate aspects in need of improvement. That mistrust and fear will likely result in a lack of candor and sincerity when completing the ratings. Unless institutions convince instructors that the primary purpose of ratings is to improve teaching and learning and will play only a secondary role in promotion, tenure, and rehiring, many instructors will continue to mistrust the process. Consequently, they will engage either in activities they think will produce higher ratings even though activities such as becoming easy graders or pandering to student consumerism may not be in the best educational interest of the students. Or they may play on student sympathies by alerting them that their job retention may hinge on high student ratings. Such behaviors, both on the part of students and instructors, characterize what can occur in a climate and/or a culture of fear. It could well be that SRTs are less trustworthy when students know that the SRTs are for tenure and promotion and more trustworthy when students know that they are used for improving teaching and learning.

Because the results of SRTs may be trustworthy or not depending on the climate of the classroom, perhaps SRTs should include some sort of climate survey to gauge the level of fear/trust that students may be feeling at the time of evaluations. With a means of determining the climate, one might have a better sense of why students have withheld comments. However, in the context of using SRTs for improving teaching and learning where students know that this is the purpose and that something will happen as a result and that the instructor/administration will listen to them and value what they say, they will probably be candid and sincere, making the results of SRTs more trustworthy in that context.

Findings Related to Students and Discriminant Validity

As discussed in chapter 2, studies related to discriminant validity concern themselves with trying to identify potential biases that occur as part of student ratings of teaching. Despite the fact that the focus of my study is on what students consider when completing their ratings and is not directly related to particular biases that may or may not occur, the results reflect a relationship to some of the biases discussed in some of the studies presented in that chapter. The following discussion parallels the three primary categories of course-related, student-related, and instructor-related biases within the section on discriminant validity in chapter 2.

Course-Related Biases.

Within the category of course-related variables, only workload/ difficulty and factors related to administration and stated purpose of the ratings apply to my findings. None of my participants mentioned any of the other factors in regard to anything that they consider while completing SRTs. These factors include the level of the course, class size, and academic discipline. I have already discussed the workload factor above, but I will comment on the

issues of administration and stated purpose of the ratings as well as class size in the context of the role that administrators can take in promoting a culture of trust.

Purpose of the Ratings. Annie, Sara, John, Diane, Andrew, and Jamie all expressed doubts about knowing what the purpose of the ratings is and/or how the process works after the students complete them. Feldman (1979) reported that ratings tend to be higher when ratings are intended for administrative purposes rather than for faculty feedback or for research purposes. Research on the effects of the purpose of rating in relation to halo effect, as Murphy, Jako, and Anhalt (1993) report, suggests that ratings tend to be inflated. My findings confirm what Feldman reports as well as provide some insight into why higher ratings occur. When the participants lack an understanding of the purpose of the ratings as well as what happens with the results, they sometimes react in ways that indicate fear and mistrust about how the administration will use the results. I discovered that higher ratings for instructors resulted from a fear that the administration will discipline or not rehire an instructor whom the participants like, or, in Andrew's case, the fear is that the administration will cancel a class or deny credit, thus affecting students. Fearing the unknown is human nature. For some participants, other emotions may combine with a fear that the instructor will not be rehired such as a feeling of sympathy (wanting to give new instructors a break), as was true for Miles, or of guilt (not wanting to be responsible for someone losing his or her job), as was true for Susan. As someone who has taught, Miles is an exception to the typical student since most students lack that experience; however, his response of overrating new instructors illustrates that students could have a variety of reasons for overrating an instructor beyond simply liking the person. It stands to reason, however, that regardless of the reason the participants give, the implicit message is that the instructor has some perceived flaws and

that the students who are overrating feel a need to conceal those flaws with higher ratings based on an underlying fear that the administration will take some sort of negative action against the instructor. Of course, a variety of other emotions could also contribute to the overrating, such as liking the instructor, being kind-hearted, and so forth, but fear of a negative administrative response is also a logical factor to include as a possible motive.

In Annie's case, the fear of an administration reaction that would somehow hurt the instructor caused her to overrate him. Based on her account, I conclude that the instructor probably fostered a climate of trust at the micro level, but the college had not done enough, at least with Annie, to create a culture of trust at the macro level. Indeed, the college that Annie attended uses student ratings of instructors as part of the evaluation process for granting tenure. When students understand that to be the purpose, coupled with the fact that the institution does not explain how it uses the results, students assume that any negative criticism could hurt the instructor. Therefore, when students like the instructor, as in Annie's example, they are more likely to withhold information and/or to overrate instructors to protect him or her. The sort of higher ratings that Feldman reported thus could be the consequence of student fears and mistrust about how the administration will use those ratings. In general, a fear of a negative administrative reaction leads to students feeling protective of the instructor or protective of themselves, but either way, the result is a lack of candor and sincerity when completing the ratings, thereby accounting for Feldman's findings.

My findings support Dwinell and Higbee (1993), who contend that what an instructor tells students about the purpose of the ratings may affect how students respond. They advocate having instructors read a standardized explanation, including how instructors and the institution will use the ratings. Providing students with information about the process

will be one step toward establishing a culture of trust. If instructors and institutions make clear how they are using the results to improve teaching and learning, the likelihood of creating trust increases. Informing students about the purpose of the ratings is critical. Providing this knowledge empowers students.

Class Size in Relation to Student Contact and Power. Although none of my participants remarked about class size, one of the studies in the review of literature appears relevant to the findings of my study. In their study related to gender bias, Baker and Copp (1997) concluded that class size impacted student perceptions. Baker found that she had better individual contact with her students teaching a smaller summer course as opposed to a large lecture class, which, in turn, they attributed to helping minimize an authoritarian professor role.

In light of my results, I find their use of the phrase “authoritarian professor role” significant. One way of interpreting this comment is that either professors behave in an authoritarian manner or that students perceive them as being authoritarian or both. As I discussed in chapter 4 in regard to power, most students enter the classroom already conditioned to view the instructor as being in an authoritarian role. Baker and Copp asserted that the perception of the authoritarian role occurs in larger lecture classes, with the assumption being that large class size enhances the perception that the instructors are authoritarian. Although class size may limit the contact with students, the factor that appears to represent the catalyst is the personal contact, not the class size.

Based on my findings, both communication and treatment of students have significant impacts on student reactions to their instructors. Through their contact with students, instructors can control the image they present. When instructors are “student friendly,” to use John’s term, and when they come across “very nice, very kind,” to use Annie’s description,

they will minimize students' perceiving them as authoritarian. Baker and Copp use the word "authoritarian" in passing, neither defining it nor elaborating on the instructor behavior that would result in students perceiving an authoritarian instructor. So at most, I can only speculate about what behaviors cause this perception, but some of these behaviors could include some or all of the negative instructor behaviors that emerged from remarks that my participants made. According to my results, these behaviors can include failing to communicate well with students both in and outside class, which includes not listening to students; conveying an attitude of not caring about teaching, lacking enthusiasm while teaching, failing to generate any interest in the subject matter for students, and not showing any understanding or compassion for legitimate student problems; and treating students inequitably or as empty vessels to fill with knowledge as well as causing them to feel disrespected and "subordinate." Even poor pedagogical practices may give students a sense that instructors do not care. Also, some instructors have strict class policies that they enforce by employing either force or authority. If instructors engage in these types of actions, they will likely convey an aloofness that students may interpret as being authoritarian. Although class size may make it more difficult for instructors to establish the type of rapport that would diminish any student perceptions of authoritarianism, it seems more likely that behavior contributes more than class size.

The results of my study indicate that course-related biases do not appear to any large extent as factors that influence student candor and sincerity. Student-related biases appeared more frequently in my findings.

Student-Related Biases.

Of those factors I discussed in chapter 2 concerning student-related biases, my

participants provided no comments related to motivation, level of the student, instructor gender, or retribution. However, prior subject interest, reason for taking the course, consumerism, feelings of indebtedness, apathy, instructor reputation, and halo effect are all biases that connect with comments my participants made.

Prior Subject Interest and Reason for Taking the Course. Research shows that students generally rate general education courses lower than electives. Marsh (1980, 1983) presents findings that favorable student ratings result from increased student interest in the subject, which apparently generates a more favorable learning environment and enhances effective teaching. The findings of Doyle and Whitely (1974) indicate that students' liking the subject correlates positively with ratings of how much the students learned, the instructor's motivational effectiveness, and overall course and instructor effectiveness. Because the factors all correlate with student learning, it casts doubt about prior subject interest being a biasing factor.

Related to prior subject interest is the reason for taking the course, which I already discussed above. Reporting mixed results, Aleamoni (1987) asserted that higher ratings resulting from students taking a course as an elective rather than a requirement is a myth. Furthermore, Freedman and Stumpf (1978) found no significance between student ratings and the reason for taking the course.

As mentioned earlier, Greg was the only participant who provided comments related to interest. He actually holds those classes he chooses to take because of personal interest to a higher degree of criticism than his general education courses. His remarks indicate that prior subject interest may cause him to hold those courses to a higher standard of quality, but they do not indicate that he will rate instructors of these courses more highly than instructors of

general education classes. Therefore, my findings appear to support the literature that shows neither prior subject interest nor reason for taking the course to be a biasing factor.

Consumerism. Despite claims in recent essays by Bellah (1999), Delucchi (2000), Delucchi and Smith (1997), Flacks and Thomas (1998), Long and Lake (1996), and Martinson (2000) that a consumerism mentality biases student ratings, no studies provide any clear evidence proving that higher student ratings result simply from instructors becoming entertainers or lowering their grading standards to pander to students. It is just as conceivable that what some may perceive as indulging student consumerism may actually reflect qualities that lead to more effective student learning and, in turn, to higher student ratings.

Nothing that the participants in my study said would indicate a consumerism mentality. As I have already discussed, some students seek dialogue with instructors. Andrew provided an example of an instructor who sought and acted on student suggestions to improve. One might argue that such behavior demonstrates consumerism. Diane said that she considers the enthusiasm that an instructor brings to the class. Does enthusiasm mean that an instructor must be an entertainer? That is not the sense that Diane gave me. While Asia's comments about students who are looking for easy graders might imply that those desires indicate a consumer mentality that might pressure some instructors to lower their standards in hopes of obtaining higher ratings, it seems just as likely that this student mentality will play a greater role in the selection of the courses to begin with than it plays a role in pressuring instructors, once the students are in a class, to give easier grades. My results provide no support for consumerism as an attitude that impacts students themselves as they complete SRTs.

Feelings of Indebtedness. In her 1995 study, El-Hassan discovered that students give higher ratings to more experienced instructors. However, she was inconclusive as to the

cause(s), wondering whether these ratings result from higher learning with more experienced teachers, from student beliefs that more experienced instructors are better, or from a feeling of indebtedness. If students reward instructors with higher ratings because of indebtedness, then this factor demonstrates a clear bias.

Annie was one participant who admitted to feeling indebted to two instructors to whom she had also given ratings higher than deserved. In her case, the primary motivation for her overrating these instructors was to protect them from a suspected negative administrative reaction:

I think you know another thing is two of those teachers that, that I did survey wrote they both wrote, uh, college recommendations for me [laughs]. I forgot about that. But I mean I, it was like I felt I owed them something. I, I was kind of in that kind of position at that time...but they were good teachers...and they should be teaching. And they're doing what they love, a-and they should stay here. So I...I, I feel... I...you know...I, um, rated them I think honestly, maybe a little...praised them a little higher maybe, just because...um...I liked them.

Based on what she said, indebtedness appears as a contributing factor for her, but this feeling of indebtedness is something she recalls as an afterthought, and it appears not as important as her caring about these instructors. Of all the participants, Annie is the only one to indicate any feelings of indebtedness. Although in Annie's case indebtedness appears to be a biasing factor, liking the instructors and finding one of the instructors to be nice and kind confound the reasons for the overrating. Indebtedness appears, therefore, to be minor but contributing factor for one student, and it is doubtful that a feeling of indebtedness is a widespread factor among students in general.

Apathy. My findings support the findings that Spencer and Schmelkin (2002) report about student uncertainty about whether their opinions matter. The authors assert that as a result, "they [students] may not pay much attention to them [ratings]" (p. 406). Greg, Diane,

and Susan indicated doubts about how much attention anyone pays to the ratings, thus supporting Spencer and Schmelkin's results. In particular, Greg displayed a general sense of apathy about completing SRTs as he admitted that he usually does not take the time to write comments. A fear that the administrators will do nothing leads to apathy and reluctance to put much time or effort into student ratings. In this respect, apathy appears to be the byproduct of a system that engenders a feeling among students that no one is listening to or valuing their opinions. Eventually, students develop the attitude that Greg displayed or that Sara reported concerning students who say, "Why bother?" because "you may not get anything done." These findings reinforce the importance of communicating the purpose of ratings, of making it clear the impact that ratings have on improving teaching and learning, and of providing students with a sense that their ratings matter.

Reputation. According to Wigington, Tollefson, and Rodriguez (1989) the highest mean rating appears in connection with the relationship between reputation and rank. They found that the highest ratings went to full professors for whom students report that reputation is a factor in their decision to take the course. Conversely, the lowest mean rating occurred with full professors whose reputation is not a factor in course choice. The authors conclude that those instructors who have developed positive reputations as effective teachers earn higher ratings and that these consistently higher ratings based on faculty reputation reflect well-developed teaching skills. In other words, a positive reputation is a byproduct of being an effective instructor, and therefore reputation is not a bias. Although my participants did not mention reputation in the context of rating instructors in line with their reputations, they did discuss the fact that students talk about instructors and that instructors develop reputations.

As I have already discussed in this chapter, Asia remarked about instructors developing reputations as being demanding (but effective in terms of student learning) versus lenient. Andrew also referred to students who “gossip” about instructors. Their comments lead to the conclusion that instructor reputations are not confined to just the effective instructors. My findings support those of Wigington et al., to the extent that it appears logical that instructors having reputations as effective instructors probably attract students who want to learn, and subsequently these instructors receive higher ratings from students. Based on the comments of these two participants, it seems more likely that instructor reputations are both positive and negative, and these reputations create certain expectations among students. However, instructors have the power to alter those preconceptions by how they conduct the class and how they treat students. Students are likely predisposed to expect a climate of fear in classes whose instructors have reputations for asserting authority and control, and it seems likely that students discuss among themselves instructors who create such a climate. In such a climate, students are likely to fear some form of retaliation whether that fear is justified or not.

Halo Effect. Overall, research studies on halo effect indicate that it may or may not constitute a serious factor creating bias in student ratings of teaching. A lack of agreement on common conceptual and operational definitions of halo has resulted in some reviews questioning the reliability of halo measures. My study is not designed to support or refute halo effect, yet some of my findings appear to substantiate some of the ideas presented as part of research results related to halo effect.

The closest that any of my participants came to indicating something approaching halo effect was Greg when he claimed that for instructors of general education courses his ratings tend to be centralized toward neutral ratings or agree or disagree rather than toward strongly

agree or disagree. His response indicates that he tends to generalize his responses for general education courses, but he also does not display this same tendency to generalize about courses that he has elected to take. The concept of impression formation that Fiske and Neuberg (1990) discussed offers some interesting insights into what may be happening with Greg and others.

According to Fiske and Neuberg, impression formation functions on a continuum ranging from category-based impressions, i.e., stereotyping, to fully individuating impressions using attribute-oriented processes. They investigated the premise that people will first attempt to fit their impressions into an existing category, but when characteristics do not easily fit into a category, people will move toward more attribute-oriented processes. Halo effect will more likely occur with students who have already categorized or stereotyped the instructor. In Greg's case, he appears to have already categorized all general education courses (or courses he is required to take) into a stereotyped preconception, and he has already generally predetermined that his ratings will fall in a neutral range.

This concept of impression formation also helps elucidate some of my other findings. As Asia discussed, students talk about instructors and their reputations. Asia's comments about how students categorize instructors into demanding or lenient demonstrate that students have an informal system of stereotyping or categorizing instructors. My contention is that the education system has also acculturated students over the years to accept (and expect) instructors in an authoritarian role, which creates another type of stereotype. Just as Fiske and Neuberg discuss impression formation being on a continuum ranging from stereotyping to individual attribute formation, I hypothesize that students enter a classroom on a continuum between fear and trust that parallels the impression formation continuum. It

seems probable, then, that students enter a classroom with varying degrees of stereotypical expectations based on either instructor reputation or acculturation along with their past experience with teachers and educational institutions. As a result, some students will enter class quite fearful because of negative past experiences involving instructors who asserted authority and control in the classroom. They are conditioned to sit quietly, take notes, and try to determine what it is that the instructor wants to hear or read to obtain a good grade. Other students may be more willing to participate if they have had instructors who have encouraged them to express themselves. These students may still be somewhat fearful but will be willing to test the nature of the class by expressing their opinions or by questioning the instructor. Depending on the instructor response, these students may participate more if they trust that it is safe to do so, or they will withdraw if the instructors demonstrate authority and control by ignoring student questions (Miles), cut off student responses when they do not agree with the responses (Annie), fail to respect student ideas (Sara), belittle students (Susan), or make sarcastic remarks (Margaret). Then one or two students may not care what the instructor thinks and will present their views whether the instructor likes them or not. In a truly authoritarian class structure, these students eventually receive low grades, which, in turn, may cause them to keep their views to themselves or to withdraw from the course. Because students are on a continuum, it would be difficult to say where they are upon initially entering a class.

If instructor behavior matches these stereotypical expectations, students will respond accordingly, and it could well be that student ratings will reflect halo effect. I contend that students expect an authoritarian environment, and because of that they can enter a classroom with a certain degree of fear. When instructors then fulfill those stereotypical expectations

through their actions and treatment of students, the fears that some students were predisposed to become reinforced. The consequence is that the instructors reinforce a climate of fear among those students already expecting that as well as establish a climate of fear among others who may have not been fearful to begin with. Conversely, when instructors act in ways that alter the stereotype by empowering students, they will begin moving students away from stereotypes and toward individual attributes with the shift being in degrees along the continuum, depending on at which point on the continuum students were to begin with. The more instructors do throughout the semester to reinforce this empowerment, the greater the likelihood that they will establish a climate of trust.

The model that Fiske and Neuberg (1990) offer also asserts that paying more attention can mitigate the stereotyping, and as people pay more attention, they will move toward particular attribute information. Additionally, they assert that motivation, coming from a third party, the target, or the perceiver's self, influences impression formation. These concepts support my contention that when instructors act in ways that break the stereotype that most instructors want authority and control, students will move from fear on a continuum toward trust and toward individual attributes of instructors. Furthermore, students will acquire more of a motive for completing ratings when they develop a clear sense of the importance of student ratings, either from an administrator or other third person who explains the purpose or from the instructor (the target). Consequently, they will take more time, pay closer attention to each question, and provide ratings based more on instructor attributes, thus reducing the possibility of halo effect. In general, Fiske and Neuberg's model implies that halo effect may or may not occur for any given student or within any given class, depending on a variety of factors; in other words, bias may or may not occur. Their model supports my

contention that by informing students of the purpose of student ratings, by acting on the results, and by generally creating both a climate and a culture of trust, students will respond by taking the process seriously, considering instructor attributes rather than stereotyping.

Instructor-Related Biases.

Those instructor-related biases presented in chapter 2 that are unrelated to my findings include grading leniency/expected grade, rank, years of experience, and type of instruction. Remarks by participants indicate that instructor personality, instructor likeability (rapport), and enthusiasm can play a part in how students rate faculty. However, in some cases it is not clear whether these are true biases or characteristics that are part of being effective instructors.

Personality/Likeability. Feldman (1986) presented results showing that positive self-regard, energy, enthusiasm, and positive regard for others had a positive association with instructor effectiveness. “Those who would contend that college teachers’ effectiveness (at least as perceived by their students) is associated with, and presumably even determine by, the personality characteristics of these teachers do not receive a great deal of support from research in which personality traits are measured using self-report personality inventories, direct self-descriptions by teachers, and related means of assessing personality traits” (p. 152). Feldman offers three interpretations of there being any association between personality traits of instructors and overall ratings of teaching effectiveness: that the results do demonstrate that personality characteristics are indeed related to their perceived effectiveness; that the measures of personality and teaching effectiveness are contaminated due to the use of the same raters for both measures; or that the results say something about the teacher’s personality, but only as students perceive it, no more or no less. Feldman

(1986) also concluded that the only positive correlations in regard to personality and ratings occur with positive self-regard and self-esteem and with energy and enthusiasm. Although Feldman's results show a relation between student ratings and instructor personality, he concludes that it is unclear whether these results indicate a valid source of influence or a source of bias. Furthermore, little appears in the research he analyzes to support contentions that college instructors' effectiveness as perceived by students has any association with personality characteristics.

Then in a later meta-analysis, Feldman (1990) concludes that although all the factors of the instructor's personality, friendliness, and respect or concern for students correlate with student achievement, these are less strongly correlated factors than 12 other instructional dimensions. Additionally, in the final analysis the dimension of instructor personality is not included as one of the 6 dimensions that explain at least 10% of the variance in student achievement, those being preparation and organization, clarity and understandableness, stimulation of interest, high standards and motivation of students, encouragement of discussion and openness to opinions of others, and elocutionary skills.

Both Diane and Sara commented on instructor enthusiasm as a factor they consider in rating factor, a finding which appears to confirm Feldman's first interpretation that personality characteristics relate to perceived effectiveness. However, no participant indicated that likeability, in and of itself, leads to a higher rating. Still, my findings demonstrate that likeability can create a bias under certain conditions.

Likeability will probably result in overrating an instructor when combined with student fears of some sort of negative administrative reaction. However, participants discussed their willingness to include constructive criticism in the comments of well-liked instructors to help

them improve their teaching. These results appear to support Feldman's third interpretation that the results say something about the teacher's personality, but only as students perceive it, no more or no less. The critical factor is whether they can trust that there will be no negative repercussions for either the instructors or for themselves. Also, students appear more ready to be candid in situations in which the instructor comes across as genuinely seeking student comments as well as when the instructor actually acts on student suggestions to improve the course. Therefore, although likeability of the instructor and enthusiasm may contribute to student bias appearing in ratings under certain conditions, these factors are just as likely not to be a bias under different conditions. My findings tend to support Aleamoni (1987), who found that students can praise an instructor's warmth, friendliness, and humor in the classroom yet still criticize the same instructor for being disorganized or for using teaching methods not conducive to student learning. However, my caveat to his findings is that this will be true only in a culture of trust, not necessarily in a culture of fear.

Enthusiasm. Closely associated with personality and likeability is the quality of instructor enthusiasm. I have already referred to Diane and Sara's mention of enthusiasm above in terms of its relationship to likeability. In the literature, most discussion of enthusiasm is in context of the Dr. Fox effect in which enthusiastic but less informative presentations receive the same high ratings that enthusiastic and informative presentations receive (Cruse, 1987). None of the comments from my participants confirm or negate the presence of the Dr. Fox effect or that instructor enthusiasm is necessarily a biasing factor.

In analyzing students' written comments to open-ended questions about the instructor, Braskamp, Ory, and Pieper (1981) found that student comments more often pertain to the instructor's knowledge of the field and ability to communicate than to conscientiousness or

enthusiasm. Their findings raise a question concerning whether it is possible the instructor enthusiasm may relate to student perceptions of an instructor's communication skills and that perhaps enthusiasm becomes confounded with these skills. My findings indicate that enthusiasm is but one of various factors students mention, and, in fact, communication appears more frequently in my participants' comments than enthusiasm, leading me to believe that while enthusiasm may contribute to higher ratings for some students, it may just as easily be a characteristic of an effective instructor. Interestingly, Dagmar's comments combined both enthusiasm and knowledge as being important to her. She associated lack of enthusiasm with older instructors who also do not stay current in their field. Instructors who do not stay current in their field are likely part of the ones who appear to be teaching just to collect a paycheck. One can imagine that seeing long-time tenured faculty that "you can't do a thing about" must contribute in some way to a climate of fear that nothing will change.

Marsh and Ware (1982) present the idea that the impact that enthusiasm may have on ratings can vary. They point out that humor, voice inflections, and enthusiasm are important factors for students who have low extrinsic motivation but are less important for already motivated students. For Feldman (1976b), enthusiasm for the subject matter is but one of several factors associated with superior college teaching. In a later meta-analysis, Feldman (1990) concludes that although enthusiasm is one dimension of instruction correlating with student achievement, other dimensions are more strongly correlated with effective instruction, including instructor preparation and organization, clarity and understandableness, instructor's pursuit and/or meeting of course objectives, student-perceived outcome or impact of the course. In fact, my findings support Feldman's results, particularly in terms of instructor preparation and organization and the instructor's pursuit and/or meeting of course

objectives as being factors about which students show greater concern than they show about enthusiasm.

Moreover, my findings align more with results that Waters, Kemp, and Pucci (1988) report. They assert that students use different criteria for the instructors they rate highly as opposed to those receiving low ratings. The highly rated instructors demonstrate personal and motivational characteristics such as enthusiasm, personality, sense of humor, and enjoyment of teaching as well as interpersonal characteristics such as helping students outside of class and learning students' names. Significantly, Waters et al., found that students did not fault instructors to whom they gave low ratings for not having these characteristics. Rather, students rate them low in terms of classroom behavior, such as "discouraging class participation, coming to class late or not at all, and lecturing without variation in class routine" (p. 204). My findings support these conclusions, both for instructors who create a climate of fear and those who create a climate of trust in terms of their approach to ratings, their communication with students, their attitude and personality traits, their treatment of students, and their pedagogical skills. The authors conclude "that 'good' faculty may be evaluated differently from 'bad' faculty and that characteristics leading to high evaluations do not necessarily have bipolar opposites leading to low evaluations" (p. 204). The characteristics that Waters et al., list for both the "good" and "bad" faculty match well with the characteristics in my findings that equate with creating a climate of trust or a climate of fear.

Findings Related to Characteristics Associated with Superior College Instructors

According to a meta-analysis of characteristics consistently associated with superior

college teachers and teaching, Feldman (1976b) identified the following as being highly valued: stimulation of interest; clarity and understandableness; knowledge of subject matter; preparation for, and organization of, the course; and enthusiasm for the subject matter and for teaching. In a later meta-analysis concerning the dimensions of instruction and their impact on student achievement, Feldman (1990) listed instructor's preparation and organization, clarity and understandableness, and stimulation of interest, which are listed above, along with the following additional dimensions: motivation of students to reach high standards, encouragement of discussion and openness to others' opinions, and elocutionary skills. Feldman contends that these characteristics consistently explain 10% or more of the variance in student achievement regardless of the method of averaging or the unit of analysis used.

Of these six dimensions, participants in my study mentioned organization, including following the syllabus, more frequently than any other factor that they consider. Asia, Jaime, Sara, and Annie all discussed organization and the importance of instructors being organized and following their syllabus. Diane mentioned "how she [the instructor] makes the students either excited or not excited about the topic that we're about to study," thus supporting the dimension of stimulation of interest. Additionally, John's description of the instructor who varies his teaching methods to appeal to different learning styles presents someone who likely stimulates student interest in learning. Finally, Dagmar, Sara, and Miles, in particular, in my study remarked about the importance of communication and establishing a dialogue between instructors/administration and students. Their comments reinforce the significance of encouragement of discussion and openness to others' opinions as one of the dimensions that impact student achievement. Despite not using the terminology, Dagmar's description of students engaging in discussions with the instructors and with the administration reflects a

basic tenant of a learner-centered classroom, engaging learners as full partners in the learning process, whereas “being told what to do” appears to describe the teacher-centered classroom and the authoritarian behavior that can lead to a climate of fear. Additionally, Dagmar’s observation that fear has a “ripple effect” throughout the institution implies that the institution could create a culture of trust through establishing a good communication process that involves everyone participating in a dialogue.

In regard to my conclusions, these characteristics of being organized, stimulating interest in the subject, and promoting discussion and openness to others’ opinions are ones that contribute to creating a climate of trust in the classroom. On the surface, pedagogical skills, such as organization, or subject-related actions, such as stimulating interest in the subject, appear not to have any direct connection to creating trust. However, when instructors are organized, follow their syllabi, stimulate interest, and engage in a host of other activities that promote student learning, they demonstrate through their actions that they care about student learning. By itself, demonstrating this type of caring may or may not create a climate of trust. If the way instructors treat students in or out of class belie that sense of caring or if they fail to promote communication, for example, or exercise authority and control in what students perceive as unreasonable ways, then the result is likely to be a climate of fear. However, effective pedagogical skills are key components that, when combined with other actions that show caring, will contribute to promoting a climate of trust. The dimension of encouragement of discussion and openness to others’ opinions, which is synonymous with the promotion of communication, constitutes a more significant element in creating a climate of trust.

Findings Related to Empowerment and Active Learning

The impact of seeking student comments early in the semester and then acting on them appears in the results of the study by Griffin and Pool (1998). Based on student responses elicited in the experimental classes, the instructor slowed the lecture rate, provided more examples of concepts discussed, and provided more hands-on activities. As a result, student learning improved and student ratings of the course and the instructor were higher at the end of the term. The instructor in the control classes did not elicit comments from the students, but the students benefited from the teaching modifications that the students in the experimental classes suggested and learned as effectively. However, student ratings of the control classes were lower. One conclusion the authors proffered is that the end-of-term rating instrument is invalid because non-instructional factors affect the results. The findings in my study would lead to another possible explanation.

Andrew described an instructor who sought student comments and acted on their suggestions to improve the class. Similarly, John described an instructor whom he considered a model for promoting active learning and teaching to multiple learning styles. These types of actions on the part of instructors provide students with a sense of empowerment. By empowering students, they become more invested in the class, which could account for the higher ratings. Both Dagmar and Sara related how student ratings involve rating themselves and their learning as much as rating the instructor. Students who have become active participants in the learning process and who have learned effectively have more of themselves invested in the course and in the instructor. If students are partially rating themselves along with the instructor, it would make sense that students who are more invested in the course personally because they have been empowered to have a say in how the instructor teaches the course will likely rate both the course and the instructor more highly

than students who have not been so empowered. The findings of Elliott in relation to attribution theory as reported by Kulik and McKeachie (1975), also support my conclusions from the standpoint that Elliott found that students tend to attribute their success to themselves and their failures to the instructor. If these findings are consistently true, it could well be that the ideas of empowerment, student investment in the class, rating themselves along with the instructor, and student success intertwine to account for the higher ratings. The point is that student empowerment is a critical component in creating a climate of trust in the classroom.

Gillett-Karam and Pena (1992) combine some of the dimensions that Feldman (1990) presented as well as helps explain the findings of Waters et al. (1988). They list the following qualities of good teachers: well-organized, well-outlined, and well-explained expectations and course goals; extended assistance; awareness of the total student; availability to students; movement from teacher-centered, teacher-in-control environment to a student-centered, student-in-control environment. The first qualities parallel Feldman's dimension of preparation and organization. The results that Waters et al., presented illustrate what can occur when instructors move from a teacher-centered to a student-centered environment. The last items on Gillett-Karam and Pena's list reinforce my contention that one of the keys to creating a climate of trust resides in empowering students.

Discussion of the Findings That Extend Beyond the Literature

Despite hundreds of studies related to student ratings of teaching, I found none that dealt specifically with the issues of candor and sincerity among students completing the ratings. As I have already discussed at the beginning of this chapter, I discovered that given the right conditions, students will or will not be candid and/or sincere. In probing what

factors may account for how students will respond, I learned that primarily the instructor and to a lesser extent the administration can impact student responses.

Participant responses about the role of administrators ranged from Miles's belief that they really have no role to Dagmar and Sara's contentions that administrators are an integral part of the process. In effect, Dagmar and Sara both described an entire college culture that promotes and encourages honest communication, not just between students and instructors but also between students and administrators, instructors and administrators, and administrators and staff. I label this as the culture of trust. Yet neither Dagmar nor Sara sees this culture as existing. Dagmar characterized students as being in an "oppressed position," and Sara referred to students being treated as "subordinates" and also used the term "oppressive" in connection with fear of being punished for negative comments. By implication, then, a culture opposite of the one they would like to see develop must exist—the culture of fear.

I hypothesize that it is just such a culture of fear that causes students not only to lack candor and sincerity when completing SRTs but also to come into the class already possessing certain other negative attitudes, such as being apathetic and taking the course simply for a grade rather than for the learning. Because this study was not intended to investigate the culture of a college, any in-depth discussion of this topic is beyond the scope of this work. However, the few comments that emerged indicate that such a culture likely exists and may assert a substantial influence on students in connection with how they eventually approach the entire process of student ratings. Ultimately, although administrators clearly play a role in the creation of the culture of the institution at the macro level, the instructor contributes far more directly to the creation of the climate in the classroom at the

micro level.

Fear is the term that best captures the unifying idea that emerged from examining the factors that led to a lack of candor and sincerity. Fear manifested itself in a variety of ways. Some participants expressed a fear that arises from a lack of knowledge and/or understanding of the rating process. The label that I created for fears that fall into this general category is “fear of the unknown.” Participants expressed doubts about knowing what the purpose of the ratings is or how the process works after the students complete them. Also, in this category are the fears that arise from not knowing what will be done with the ratings or what difference the impact will be on the instructor if one responds with “strongly disagree” as opposed to “disagree,” for example. Not understanding the process and how it works feeds other fears, particularly the fear of retaliation and the fear that nothing will change even if students indicate that problems exist.

Fear of retaliation occurs both at the instructor level and at the administrative level. The primary fear the participants expressed concerns retaliation of some sort. Some participants expect the instructor will direct this retaliation toward them individually because of the instructor’s ability to identify them somehow (by being present in the room and seeing them writing, by asking students to put their name on comments so that they can “earn extra credit,” by knowing a student’s handwriting, by associating the nature of the comment with a particular student, or by the phrasing a student uses). Or they expect that the instructor will punish the class as a whole by making the class more demanding. In particular based on my findings, I hypothesize that grade-conscious students appear to be more sensitive to fear of retaliation and may be even less inclined to write anything negative for fear the instructor will connect the comment with them and will lower their grade. What surprised me is how

students who claim that they are candid and sincere in most circumstances have experienced situations that prevented them from writing a word because of their extreme fear of being identified. Fear appears to have a much greater impact on student ratings than anything I found in the literature would indicate.

Participants mentioned a variety of other factors connected with fear. Related to the fears associated with anonymity and retaliation is the issue of the timing of when the institution administers the rating. Some participants expressed concerns about the ratings coming too early during the semester, and some suggested having ratings occur at the end of the course or even after the course with the students mailing the ratings to the college. Any steps that the institution can take to ensure anonymity will help establish a sense of trust. Two ironies exist, however. The first irony is that in a climate of fear, students will still fear being identified regardless of any steps the institution may take to try to alleviate student fears. The second irony is that in a climate of trust, students may have no concerns about being identified because they know that no retaliation will occur. Therefore, ensuring anonymity is likely insufficient, by itself, for producing a climate of trust necessary to elicit candid responses from students who are truly fearful. Establishing that climate appears to require not only that instructors encourage students to be honest and to provide the security of knowing that their identities remain anonymous but also that instructors and administrators begin promoting communication with students from the outset of the class. As John pointed out, student ratings have “the potential of being a good communication device” between the student and the instructor or between the student and the administration. Realizing this potential for communication begins long before students complete an SRT.

Moreover, this situation of the timing of when students complete a rating creates a

paradox. The instructors who likely need the comments the most so that they can modify their teaching are the ones who should receive the comments early in the semester so that they can adjust; however, if students perceive some threat of being identified and of the instructor's retaliating somehow, they will likely withhold their comments. So early ratings would probably create a false impression of what the instructor is really doing (or not doing). If students complete their ratings at the end of the course, the likelihood of obtaining greater candor and sincerity increases, but how do the comments benefit those students in terms of changing instructor behavior for the class that they have just completed? Furthermore, instructors receiving low ratings may be inclined to dismiss the results, contending that it was "just a bad group of students," and will change nothing the following semester. This is a problem with which administrators and faculty members alike will need to wrestle. One solution would be to encourage instructors to conduct an informal assessment using Classroom Assessment Techniques within the first two or three weeks of the course before students have formed a real opinion about how an instructor will react (provided the instructor does not make sarcastic remarks about student ratings on the first day of class). Of course, instructors will need to demonstrate that they will respond to these comments in an open manner without retaliating against the students. Throughout the semester instructors could initiate a series of these informal assessments aimed at formative evaluation rather than the summative rating that formal SRTs provide. These formative assessments will likely generate feelings of trust among the students, which, in turn, should result in more candid and sincere ratings when students complete a formal summative rating of the course and the instructor.

Participants mentioned a few other fears but less frequently. These include a fear of

some sort of administrative reaction that will either hurt the instructor (disciplining or non-rehiring) or hurt the students (denying credit or canceling a class), or conversely they may fear no administrative action and no change on the part of the instructor will occur, so they do not take much interest in completing the ratings. The perceptions (or misperceptions) of the role of administration appear to arise because students lack information about the real purpose of student ratings and the function of administration in the process. As Wheatley (1992) asserts, “If information is not available, people make it up. Rumors proliferate, things get out of hand—all because people lack the real thing. Given the need for constant nourishing information, it is no wonder that ‘poor communication’ inevitably appears so high on the problems list” (p. 107).

Andrew’s fears about the administration canceling classes if they learn that an instructor has missed too many classes demonstrates just how far students can and do invent consequences. Personally, I have never heard of a college canceling a course or denying students credit in a situation like this, and Andrew did not say he had ever experienced this happening. However, the fact that he imagined such an administrative response demonstrates just how intense feelings of fear can become. Andrew’s reaction of just not saying anything—a “better safe than sorry” attitude—could well typify how some students approach rating some courses and instructors. Yet if the comment section is blank, how would the administration or the instructor, for that matter, know what is really happening in the classroom? What could be most troubling about students like Andrew would be if the class contained a majority like him (although perhaps not quite as paranoid) who kept their ratings on Likert-scale questions in the middle and who wrote nothing. If the majority did that, the few students who chose to be candid might appear to be an anomaly. One would never

know. When fear reaches beyond the classroom and an instructor's reaction to the level of a feared administrative reaction like this, such a reaction may reveal that the climate of fear present in the classroom extends to a culture of fear within the institution itself. If enough students engage in this process and when rumors spread, as my findings illustrate is happening, it is not surprising that the result is a lack of candor and sincerity from some students.

Another fear comes because of other students who may under- or overrate an instructor causing other students to slant their ratings in the opposite direction to counterbalance the other students' ratings. Interestingly, none of the participants indicated that they themselves had underrated or would underrate an instructor because of receiving a grade lower than they believe they deserve. Nor did they say that they know with certainty that students with lower grades are intentionally underrating instructors. One might question whether no one admitted to engaging in underrating because of the possible embarrassment or guilt of revealing such behavior, and it could certainly be true that the participants may have been less than candid about underrating instructors. However, given the overall high grade point averages of the participants, it seems unlikely that they would have any motive for underrating faculty as they clearly do not fall into the category of the student receiving a low grade. Yet intentional underrating is certainly an action that students suspect other students of participating in. So it is difficult to know to what extent intentional underrating occurs if at all based on my findings; however, both students and instructors fear that such underrating occurs. Only John indicated that he might underrate an ineffective instructor if he knew that other students were planning to overrate that instructor, but he did not admit to ever underrating an instructor. Participants also mentioned that students talk about instructors and how they have rated

instructors outside of class. Although the participants did not specifically mention discussing fears about instructors, it is easy to imagine that when the discussion is over who is “easy” and who is “hard,” some students will become fearful of taking hard instructors, and it is also easy to imagine that students discuss how instructors treat students.

Two of the participants expressed fears related to how instructors might perceive them personally as well as related to feelings of guilt. Andrew avoided making comments that might be perceived as being politically incorrect by expressing a view that too much time was being spent on ethnic diversity. Susan avoided making comments about not being able to understand instructors who were non-native speakers whose English she could barely understand. In her case she did not want to appear biased. This type of fear, however, appears to be the exception rather than the norm, but these examples demonstrate that fear appears in various forms, but whatever the fear, it impacts student candor and sincerity. The difficulty, of course, is that controlling the sources of student fears or potential student fears may be impossible. One would hope, however, that in an environment in which students learn to think critically, as Sara suggested, as well as an environment that emphasizes learning rather than grades and when students trust that purpose of their ratings is to improve teaching rather than to cause an instructor to lose his or her job, students will be more candid and sincere.

Although uncertainty about what purpose the ratings serve and what the administration does or may do with the results causes a certain degree of fear, the primary source of fear in the classroom is the instructor. Besides the fears that arise because of the unknown or a general fear of retaliation, participants have also experienced firsthand negative instructor reactions, coming in the form of sarcastic comments about prior student comments or by

increasing student workloads. Participants mentioned other specific instructor behaviors that appear to contribute to creating a general climate of fear or of trust in the classroom. I classified their actions into five categories: actions in relation to the ratings, communication skills, attitude/personality traits, treatment of students, and pedagogical skills and practices.

Participants commented on specific actions that instructors have taken or could take that will cause them to be less or more likely to be candid and sincere in their ratings. Instructors who make sarcastic comments about prior student ratings are almost certainly going to cause some students to fear expressing any negative criticisms, and instructors who comment about the ratings having an impact on their future employment can evoke student sympathy resulting in overrating. Additionally, Susan indicated that the instructor's presence in the room during the ratings would cause her to withhold comments. An instructor's being in the room serves to remind students, perhaps subconsciously, who has the authority and the ultimate power in the class so that if students have any doubts about whether the instructor will retaliate because of negative comments, just seeing the instructor could cause some students to keep their comments to themselves.

Andrew spoke of his experience with an instructor who encouraged student honesty, and other participants provided suggestions about actions that both instructors and administrators could take that would promote more candor and sincerity when students complete SRTs. The success of these measures for achieving this candor and sincerity appear bound, to a great extent, on other factors, such as the instructor's ability to communicate effectively with students, his or her behavior throughout the course that reflects a positive attitude toward teaching and toward students, and his or her treatment of students. It is unlikely that just having an instructor tell students that they should be candid

and forthright in completing their ratings will, by itself, necessarily elicit the desired student response. Without the other conditions already being in place, conditions that promote student trust in the instructor, students will still likely withhold their comments or overrate the instructor. Recall Dagmar's example described earlier in which her instructor asked students at the end of the course to write an evaluation of what they had learned in the course. Even the reward of 10 bonus points could not entice Dagmar to write an evaluation. However, situations do exist in which instructors had established enough rapport with students to cause them to respond openly and candidly.

As I have already discussed, just requesting student honesty is no guarantee, in and of itself, of promoting more candor; however, participant comments indicated that when instructors explain the purpose of the ratings and ask for honesty, students are more likely to respond accordingly. Instructors and/or administrators could enhance the chances of developing a climate or culture of trust by explaining the purpose of the ratings, by instructors' discussing how they have used the ratings to effect improvements, and by demonstrating that they are enacting student suggestions as a result of soliciting student comments early in the semester. In other words, it helps students to know that their comments do matter and that people are listening.

Critical to the success of creating trust but a factor that can just as easily create fear involves communication in general, not merely in connection with what instructors say about student ratings or the rating process. Communication plays a key role in creating a climate of fear or trust in the classroom. According to Wheatley and Kellner-Rogers (1996), "We do know that in healthy human systems people support one another with information and nurture one another with trust. Our wonderful abilities to self-organize are encouraged by openness"

(p. 39). When instructors fail to engage in a dialogue with students or when they fail to answer student questions or cut off student answers when they disagree with the student or appear not to like the students' answers, they send the message to students that they do not value or want to hear what the students have to say. In a case such as the one Miles cited in which the instructor would not answer student questions, the climate of fear developed so early in the course that students actually withdrew. It is possible that some of the students who withdrew did so because they felt they would not earn an A from this instructor, but if students were withdrawing before completing any exams or graded assignments, it is more likely that they withdrew because of the classroom climate. The climate that this instructor created could well have led to a lack of candor and/or sincerity on SRTs for some students, but that does not even count the students who had withdrawn because of the climate before they had ever reached the point of completing an SRT.

Students need to feel comfortable talking with their instructors both in and outside of class as Jaime's comments indicated. Her saying "even if they're inviting to that" is telling. Just announcing office hours or availability outside of class will not guarantee that students will accept the offer. A student like Jamie needs to feel comfortable with an instructor, and the classroom climate will likely have a large bearing on that comfort level. How instructors communicate with students in class likely influences the comfort level as well as conveys their attitude toward students. Participants made it clear that establishing good communication between instructors and students as well as between administrators and students/instructors provides an essential component for establishing a climate of trust. When everyone engages in a dialogue, everyone shows respect for others' views and displays a willingness to listen. The instructor Margaret described who made sarcastic remarks about

criticisms that she had received is another example of instructors who, through their actions, send the message to students, “I don’t want to hear you.” In other words, “I don’t want to listen.” When instructors give students that impression, it is no wonder that students shut down and refuse to engage in providing candid remarks. By applying critical thinking throughout the class and by respecting students’ views enough to listen to them, instructors demonstrate that they value student opinions and ideas. Creating an atmosphere that promotes communication between the instructor and the students also is one of the ways in which instructors convey their attitudes toward students.

The participants mentioned other attitudes that appear to contribute to creating a climate of fear. These include conveying a sense of “being there” for a paycheck rather than being there for the students, lacking enthusiasm while teaching, failing to generate any interest in the subject matter for students, and failing to show any understanding or compassion for legitimate student problems. Participants mentioned a number of attitude-related factors that they consider when rating instructors. These include how instructors interact with students, their personality, and their enthusiasm/passion. Diane, for example, mentioned both the instructor’s personality and enthusiasm as factors that she considers in completing her ratings. Her comments imply that an instructor’s personality and enthusiasm are characteristics that she is looking for in her instructors, probably because these traits create a positive learning environment and lead to development of a climate of trust, not that she is seeking an entertainer. Making class fun is a factor that Diane also mentioned, but in this case she indicated how much fun the class is depends on “the teacher and her attitude.” Sara characterized the attitude of some instructors as “teaching joy.” Contrast this type of teacher with the ones that Miles described as just “being there for a paycheck.” An instructor who is

“teaching joy” is bound to inspire a climate of trust in the classroom as well. Other characteristics that reflect the instructor’s attitude and personality that participants mentioned include sincerity, honesty, diligence, and being able to display a degree of vulnerability.

When Miles spoke of instructors’ making themselves vulnerable, he said, “...making yourself vulnerable I think helps a lot ‘cause then you put yourself in the students’ boat, you know. You’re right there with them.” Of greatest significance are these statements about the instructors’ putting themselves “in the students’ boat” and being “right there with them.”

Students often feel vulnerable (Dagmar’s word was “oppressed,” and Sara’s word was “subordinate”). This vulnerability can contribute to a climate of fear when instructors conduct class in such a way to continue or to reinforce their feelings of vulnerability.

However, when instructors make themselves vulnerable through certain admissions, they can give students the impression that they are “right there with them.” Displaying a certain amount of vulnerability helps humanize instructors. Miles also seems to be saying that students are likely to mirror an instructor’s honesty. When instructors are honest with students and instructors make themselves vulnerable, students may feel safe enough to be honest, and they may make themselves vulnerable as well. Yet as Sara’s example of the instructor who came to class admitting being unprepared, making one’s self vulnerable can extend too far.

Sara saw the instructor as making excuses. Clearly, being vulnerable does not mean that instructors can admit being unprepared and unsure of how to grade because such admissions likely undermine the confidence students have in the instructor’s competence. To admit an error and to try to correct it may demonstrate the type of vulnerability that would likely create trust whereas to admit to not putting forth the effort even to prepare a syllabus

for the first class would likely create a level of mistrust. Paradoxically, the message is that instructors should make themselves vulnerable by admitting some mistakes but not making themselves vulnerable by revealing others.

By itself, attitude may or may not have much impact on the overall climate. For example, just because an instructor lacks enthusiasm while teaching will not necessarily cause students to fear being candid or sincere in their ratings of that instructor or course. Yet lack of enthusiasm could set a tone, and when attitudes combine with other factors such as communication and treatment of students, an attitude like lack of enthusiasm could help create distance between the instructor and the students. Certainly, treatment of students appears to have a more direct impact on what climate the instructor creates in the classroom.

Participants mentioned a number of factors that relate to how instructors treat students. According to the participants, these factors are ones that they consider when rating instructors. Instructors can treat students in a variety of ways that help either to cause students to feel devalued or valued. Jamie, for instance, looks for how equitably instructors treat students. Dagmar mentioned about how some instructors treat students as if they “are stupid” and as if “they need to be filled up with, you know, nonsense knowledge.” For Sara, the treatment involved being treated in a way that caused her to feel as if the instructor did not take her seriously and did not honor or respect her so that it affected her sense of self-esteem. When instructors treat students in ways that appear discriminating or belittling, causing students to feel inferior, to find that a climate of fear exists in the classroom is not surprising.

Conversely, when students, such as Annie, perceive instructors as nice and as kind, in other words as caring, they appear more likely to overlook pedagogical shortcomings,

especially if they fear any negative consequences befalling these caring instructors. In a trusting environment, students are more likely to provide recommendations for improvement to any instructors they feel are caring but will do so in an equally kind way, trying to downplay criticisms so that the flaws appear less significant than the virtues as John's comments indicated. Linked with being nice and kind are other ways that participants spoke of how instructors treat students, factors that they consider when completing SRTs. These include being, as John said, "student friendly and learning oriented" and treating students equitably and with fairness in class, according to Jaime. The comments from Annie and John illustrate the difficulty determining to what extent liking the instructor is a bias that affects rating reliability. Annie's remarks provide evidence to support that liking an instructor creates a bias that impacts reliability whereas John's provide evidence that liking the instructor may be the result of the instructor's being an effective instructor and although liking the instructor may result in John's trying to downplay what the instructor (or the administration) might perceive to be negative criticisms, he would, nonetheless, provide those comments. Which way a student reacts depends on the climate of the classroom as well as the culture of the college. Sara's remarks about instructors' giving students permission to be honest reflects yet another way that instructors treat students that can impact how they complete SRTs, but as she notes, instructors cannot expect students to be candid just on the basis of asking them to at the time they are completing their ratings. Instead, this is a practice that must occur from the beginning of the course. I take Sara's comment to mean that, in effect, "actions speak louder than words." Instructors cannot merely "tell" students that the goal is to create an atmosphere of openness; instructors must practice and promote openness and honesty. The result will be that students will be honest without the instructor even

having to ask. Treating students in such a way so as to cause them to feel “liberated” and treating them with respect will likely result in the creation of a climate of trust.

Diane mentioned instructors who fail to return work in a timely manner yet who demand that students complete the work, reflecting that they are “not as dedicated to correcting the work as [they] are to assigning it.” Although she is referring to a negative situation, one can assume that the corollary would mean that students likely view instructors who return work in a timely manner as being dedicated. It is also likely that not only do these instructors earn student respect because students perceive that they are being respected, but also students likely develop a greater sense of trust with these instructors because they see a demonstration of their dedication to student learning.

The results of this study reveal an interesting phenomenon occurring in relation to an instructor’s pedagogical skills, but this phenomenon is consistent with the general comment that I made at the beginning of this chapter that student ratings are and are not trustworthy. The phenomenon is that weak pedagogical skills may or may not result in students rating instructors appropriately. As previously pointed out in the discussion of students’ overrating instructors, participants disclosed that they have overrated instructors who had teaching flaws, such as being disorganized, because they like these instructors and fear that somehow the administration will punish these instructors by reprimanding or even firing them. Therefore, ineffective teaching practices evidently do not, in and of themselves, result in lower ratings. When students find themselves in a climate of trust, they show less reluctance to provide honest and balanced criticism to help the instructor improve. A recurring theme among some of the participants was that everyone can improve, even the good instructors. As Asia remarked, “You can always tell, let a person know there’s always room for

improvement. This is a good course or you're doing a great job or...but there is some things that you need to take consideration of when your next class comes around or your next job or your next evaluation or whatever comes around."

Closely related to pedagogical skills is the instructor's knowledge. Perhaps an instructor characteristic that students may take for granted is the instructor's knowledge. Despite the fact that knowledge is not pedagogy, it is one of the two pillars of a successful instructor, the other pillar being pedagogy and the skills to promote student learning. An instructor's knowledge provides an important component of student respect, which, in turn, is a likely contributor establishing a climate of trust provided that the other elements of trust, as discussed above, are also present. Diane listed the instructor's knowledge in the same sentence with the instructor's ability to excite students about the subject matter as a factor that she considers. Recall, too, Annie's comments about her respect for the education level that her instructors have attained and how she thinks of them as "demigods" because of their achievement. As long as instructors maintain currency in their field, unlike the example Dagmar gave of the instructor using materials from 1988, an instructor's knowledge can earn student respect. The instructor's knowledge must also combine with the instructor's skill in promoting student learning.

Although pedagogical skills and practices are likely not direct contributing factors to the creation of the classroom climate, whether an instructor is a skilled practitioner appears to reflect his or her attitude about teaching. Therefore, like attitude, when pedagogical skills and practices combine with other characteristics related to communication and treatment of students, they can add or detract from the overall climate in the classroom. Pedagogical problems that participants mentioned include being disorganized and/or not following the

syllabus; using outdated course materials; failing to prepare students adequately for exams either because of not covering some material or moving too quickly through material before students have a chance to learn it; and failing to return assignments in a timely manner. For instance, Asia expressed frustration with instructors whose classes lack organization and clarity and move too quickly through complex material. When students “get frustrated” or feel that when the instructor’s failing to help the students meet the instructor’s expectations is “unfair,” these emotions possibly lead students to perceive the instructor as uncaring. Attitude toward and treatment of students could then conclude in creating a general climate of fear. Diane’s frustration comes from instructors who demand more work from students than they are willing to grade and return in a timely manner. Although Diane did not use the term “unfair,” as Asia did, the implication of her remarks points to a practice that students could view as unfair or hypocritical. The underlying message to the student is, “I expect you to work hard, but don’t expect that from me.” Such behavior may also contribute to the feeling some students have of being in a “subordinate” role because the students may perceive the instructor as not working hard but being able to get away with that because the instructor is the person in charge. By itself, such behavior may or may not produce a climate of fear, but it could be a contributing factor depending on other conditions that concurrently exist in the classroom with a particular instructor.

However, my findings indicate that a hierarchy may exist among factors in terms of their impact on SRTs. For example, Annie overrated an instructor, despite his being disorganized, because she found him to be caring and nice. The implication, therefore, is that how instructors treat students may cause students to overlook certain pedagogical flaws. However, without other offsetting behaviors, poor pedagogical skills and practices may

signal a general lack of caring which may contribute to the apathy that Greg felt, or they may combine with explicitly negative comments or actions from the instructor to contribute to creating a climate of fear. I wonder, too, whether Greg's general apathy reflects the culture of students whose primary interest is the grade rather than the learning. Perhaps over the years various factors, such as instructor behavior, lack of communication with students, a sense that the instructor is in control, in addition to poor pedagogical skills and practices, have created this culture along with the creation of a climate of fear that students can do nothing to change the system,. Whatever the cause for Greg, his lack of interest in the class appears in his lack of interest in putting any time in on the rating to write comments.

On the other hand, the instructor's knowledge and pedagogical skills and practices also contribute to a climate of trust. Knowledge and pedagogical skills alone will not necessarily produce a sense of trust. Through their comments and their treatment of students, instructors can convey the sense that they consider the role of students to be passive receptacle of knowledge and that they do not value student comments. However, it also seems clear that a combination of knowledge and pedagogical practices that promote excitement about learning as well as the learning itself will increase the likelihood of building a climate of trust. It could also well be true that instructors who are secure in their knowledge and teaching abilities are also more inclined to empower students through the use of collaborative and active learning techniques, part of their pedagogical skills.

The results of this study indicate that when instructors present their knowledge in a way that excites student about learning, they create an atmosphere that is conducive to building a trust relationship with the students. Instructors who adjust their styles to teach to different learning styles, who treat assignments as means to promote student learning rather than

merely as a tool to assign grades, and who adjust their teaching pace to ensure that students understand so that they have the foundation needed for future learning also show through their treatment of students that they care about student learning.

Instructors' pedagogical skills and practices do not appear in the participants' comments directly in relation to creating the climate in the classroom. However, an instructor's teaching skills may contribute to other factors, such as signaling the instructor's attitude toward the students or teaching in general or displaying a treatment of students. These other factors, then, aid in producing the climate of fear, and this climate, in turn, impacts the candor and sincerity of student ratings. Determining whether a true correlation exists between ineffective pedagogical practices and creating a climate of fear is beyond the scope of this study, but it may warrant further investigation.

More than any other factor related to pedagogy, students mentioned class organization and adherence to the syllabus as factors that they consider when rating courses. Organizational failures can undermine student confidence in an instructor's knowledge and ability to teach. Disorganization and inability to follow the syllabus create a general mistrust of instructors' competence, which may lead to a general fear about expressing that if students have any doubts about how the instructor and/or the administration will react.

Finally, instructors can create a climate of trust when they promote critical thinking from the outset of the course. By creating an environment in which students participate in applying critical thinking skills, instructors will have laid the groundwork for promoting candid and sincere comments from students. When instructors who engage students in critical thinking practices in the classroom then ask their students to apply those same skills

in completing their student ratings, the chances of receiving candid and sincere responses are greatly enhanced. Based on what Sara and Dagmar expressed, apparently a transformation of values occurs for some students when they engage in critical thinking and dialogue in the classroom. Rather than being strictly grade oriented, students may engage in a process of self-examination and self-assessment in terms of their learning. This shift in thinking could be significant when considering that fear of grade retaliation is one of the main reasons that students withhold comments. If students stop focusing on the grade but concentrate instead on the learning, a reduction in the fear of retaliation may also occur. The transformation also creates a different type of atmosphere in the classroom. Sara said, “It’s like, y-y-you create this atmosphere whether it be...uh...by raising a question...uh...like let’s say...I think I thought of this recently, um, just asking a question.” This “atmosphere” that Sara is referring to equates with what I am labeling as a climate of trust. When critical thinking is an ongoing practice in the classroom, Sara characterized it this way: “It’s, it’s so powerful, um, you know. It’s, it’s so...like, moving?”

To conclude this section, I want to add my observations about the students themselves and their attitudes to ratings in general. As mentioned at the beginning of this section, all 12 participants asserted that under the right conditions, they are candid and sincere in ratings. Several of the participants demonstrated having a sense of responsibility for being candid and sincere when completing SRTs as long as they feel safe doing so. Dagmar even discussed the rating process as one in which she rates her own learning as part of the process of rating the instructor. For Margaret, the issue is not whether instructors pay attention to the comments; she will provide suggestions for improvement even if the instructor chooses to ignore them. Yet Margaret adamantly refused to write comments on an instructor’s rating for fear of being

identified and because she knew that she would have to take the same instructor again because of the program she was in. John and Asia also indicated that they include suggestions for improvement even for instructors they think are effective. They want to be helpful and fair. In chapter four, statements from Dagmar, Margaret, Sara, and John illustrated the value of completing SRTs to them. Just from the comments of these four participants, a picture of these students emerges in terms of how they approach the process of writing comments about instructors and classes. They are serious. They concern themselves with fairness, truthfulness, balance, and objectivity. When they are not fearful, they want to provide the type of constructive criticism that will help the instructor improve. Clearly, students are willing to be candid as long as they feel empowered to do so or, as Sara put it, as long as they feel they have “permission” to do so.

Participants also put SRTs in the context of being an effective way of communicating with instructors and/or administrators. Again, to use Sara’s words, the process gives students “a voice.” As some of the participants indicated, the comment section may simply be a place for them to express their gratitude, not just a place to denigrate an instructor or a course. Jaime, in her comment about using the comment section to praise instructors, ended that comment by saying, “...what maybe they could do to change a little bit or to get through to some students more.” As the last clause indicates, however, the comment section also provides a place to qualify what she writes and to provide suggestions for improvement. Her comments appear to indicate that she actually prefers praising her instructors rather than criticizing them in a negative way, and she added the part about writing suggestions almost as if she were apologizing for having anything negative to say. It might be fair to say that a number of students may be like Jaime in this respect in wanting more to praise but also to

provide suggestions if they can.

Andrew and Annie even showed great respect for instructors, saying that teaching is “an honorable profession” and that professors are like “demigods.” If anything, the reverence and respect that these participants give the profession of teaching demonstrates that at least for some the likelihood of positive ratings outweighs the likelihood of negative ones. Susan commented about trying to remain objective and separate her personal feelings about an instructor from that instructor’s ability to teach. Such comments indicate that, if anything, students lean toward wanting to be positive in their ratings but that they will be willing to provide suggestions for improvement as long as they trust the instructor and the process. Trust appears to be the key ingredient for eliciting candid and sincere ratings.

Based on the comments of some of the participants, it is clear that at least some students enter their classes with certain preconceptions about their responsibilities when it comes to providing a critical—both positive and negative—rating of a course and its instructor. The comments that indicate students feel a sense of shared responsibility, that they are rating themselves as well as the instructor, along with statements about being honest and not intentionally lying to (false flattery) or about (underrating) an instructor all provide examples of preconceptions that students bring with them when entering the classroom. I would contend that most, if not all, students enter a course generally open to being honest. Some instructors, for example, use a Classroom Assessment Technique of having students write about their expectations of the course, their prior knowledge, their fears about the course, and so forth on the first day of class. These instructors usually explain to the students that their purpose is to gain a sense of who they are and what they are expecting to learn. Having used this technique myself when I taught English composition, I know from

experience that student responses were candid and frank. They admitted to receiving only “C’s” or “D’s” in high school, of having writing phobias, and of taking the course only because they “had to.” Eliciting comments like this before students have a chance to form any opinions about whether they should fear or should trust an instructor almost always results in openness from students. The point is that students generally enter class probably more willing than not to be candid and sincere. With students who have experienced completing an SRT before, some also enter with a sense of responsibility about their role in the class as well as other preconceptions about their role in the rating process.

Asia remarked about the importance of students needing to “come across as an adult,” which means couching negative comments in the form of positive suggestions for improving the class rather than in the form of comments that simply attack the instructor. Another example of the sense of responsibility that Margaret expressed is the comment about restricting criticisms to the instruction-related issues rather than personal criticisms such as not liking the how the instructor dresses.

Susan and Greg indicated that they have established their own norms for rating instructors. One point of highlighting what are more clearly norms in what Susan and Greg presented is to show that students put some thought into the standards they use to rate instructors rather than basing their ratings simply on emotions. The rating instrument questions help to target specific areas of evaluation, but students must still apply some norm to distinguish between the gradations on the Likert scale. Whatever they use as a norm, the impression the participants gave me is that they think about their ratings for the most part.

If students are applying individual norms during the rating process, it would appear to be impossible to determine when it occurs. In the context of this study, these examples

provide evidence that students bring preconceptions about the general norms they will apply in completing ratings as well as reflect that they probably think about their ratings rather than rating instructors based merely on their mood. What is most interesting about their norms, however, is that Susan's norm compares instructors against other instructors (either current or previous) whereas Greg's compares his preconception of what he thinks he will or should learn upon enrolling for a course against what he actually learns. On the one hand, the fact that at least some students have a norm by which they rate instructors illustrates that ratings are not arbitrary and based on more than the student's mood. On the other hand, the fact that variation occurs in the standards students use points to another way in which the reliability of ratings fall into question, for one student could rate an instructor high compared to other instructors whereas as another student in the same course could rate an instructor low because the instructor fails to provide the student with the level of learning that the student had been anticipating. Such a finding indicates the importance of providing some training for students in "how" to rate instructors and to develop a somewhat common norm.

The findings also indicate that students are discriminating in terms of rating instructors according to whether they are demanding or lenient, with the deciding factor being whether the student wants to learn versus obtain an easy grade. The results of the interviews confirm what most educators already know to be true—that many students have more concern for obtaining a good grade than learning. The results also confirm the fact that students do discuss instructors and ratings of instructors among each other, resulting in certain instructors developing "reputations." However, the fact that some students do seek instructors who will teach them something, even if it means working hard, shows that students value learning. Although it is possible that those students who are searching for an easy grade may rate the

“lenient” instructors highly, it is just as possible that those students who are searching to learn something will rate the “demanding” instructors highly.

Although some of the participants expressed uncertainty about or mistrust of the rating process, participants, such as Dagmar and John, displayed a fairly sophisticated understanding of how the process works. This understanding includes the belief that administrators are likely to view an extremely negative rating that simply attacks the instructor that is out of character with the other ratings as being an anomaly and that student ratings are important and worth doing. These observations about the likely administrative reaction to negatively slanted student comments reveal a sophisticated awareness of what types of student comments will likely have little effect. John also shows an awareness that presenting balanced comments will more likely lead to the administrators’ and/or instructors’ taking them more seriously. His reaction also indicates that he believes that someone is reviewing these evaluations. Although a number of the participants, as discussed earlier, wonder if anyone reviews the ratings, John demonstrated the opposite view. Additionally, for someone like John, who believes that administrators are paying attention to the comments, his response about providing balanced comments exemplifies how seriously some students take the process. Annie even expressed disappointment about not being able to rate all her faculty. As Sara contended, the ratings give her a “voice.” These various reactions point to an inherent trust in the process itself—that it will have a positive impact.

It is not the rating process that causes trust or a lack thereof; it is the behavior of instructors and administrators that influence student candor and sincerity. So not all students are naïve about ratings, and most of my participants desire to participate. However, this finding probably says more about those students who volunteered for my study as opposed to

the student body in general. It is even likely that the majority of students may be more like Greg, who tended to be apathetic to the process.

Students talk about instructors and classes. Comments from both Andrew and Asia, cited earlier, verify that instructors have reputations and that “rumors” spread among students. Although their remarks were in the context of students’ discussing their ratings among each other after the class and discussing “hard” and “lenient” instructors, I would also assume that students discuss instructors that they find effective or those they like. Just as negative comments could precondition students to expect the worst and anticipate a climate of fear, so, too, could the positive comments cause students to enter a course expecting a positive experience and being open to trusting the instructor. In addition to this type of preconditioning, some of the participants spoke of their expectations of how an instructor should handle instruction or relate to the students.

Students like Susan, who bases her ratings on comparisons with previous instructors, may bring certain expectations about how instructors should be. “I look to teachers in the past who I’ve learned the most from, and so, you know, like approachability, for example...is really important...in my opinion if, in a professor. Or, um...you know, the, are they dynamic when they present their information?”

Throughout the interviews, participants provided a number of positive references to instructors through specific examples as well as through remarks about how they view them. I have already cited Annie’s and Andrew’s views of instructors. Sara recognized the difference in knowledge between the instructor and the students, which elevates the status of instructors. Such reverence or respect for instructors illustrates a desire among some students

to trust the instructor. These positive views balance the underlying fears that other students feel. The difficulty is that students do not fall neatly into two categories of those who trust and those who fear. Instead, students' predispositions about instructors likely fall on a continuum ranging from trust on one end to fear on the other. Those students who fall at or close to the trust end of the continuum may need only one overt act, such as an instructor making sarcastic remarks about prior student ratings, to cause them to leap to the other end of the continuum. In other cases, it may be poor communication over a period of time that could cause students to gradually shift away from trust and toward fear. For students who begin more toward to fear end of the continuum, I hypothesize that it will take them longer to develop degrees of trust in an instructor. The process of moving from fear to trust will take time unlike the leap that can occur from trust to fear. At least within the context of the classroom, the core controlling factor is the type of power an instructor exerts. The instructor's abuse or use of power will determine the extent of trust or fear students ultimately feel. To understand more fully how this occurs requires a discussion of the nature of power.

Future Research

Besides conducting more studies about the use of power and its role in effecting a classroom climate, other topics that emerged in connection with this study deserve future attention. One such area concerns student attitudes about grades. Elliott's study, as cited by Kulik and McKeachie (1975), that links students' ability levels and the level to which the instructor teaches as impacting ratings might combine with research related to attribution theory and student attitudes about grades. A connection might exist between whether students attribute their performance to their own effort (or lack of effort) or to the instructor

along with their grade earned. Related to this idea is to see if a correlation exists among the factors of attribution, grades, and empowerment. Possibly, students whom instructors empower rightly attribute successful learning performance to themselves. It could also be that grade consciousness heightens student sensitivity to instructor uses of power in the classroom, and that would be another area of possible investigation.

Also somewhat connected with the degree of concern over grades and current grade record is the relationship between such factors as prior experiences with instructors that may predispose students to be fearful or trusting along with the degree of self-esteem and self-confidence students feel when they enter a classroom. These factors, and possibly others, may determine where on the continuum between the two poles of fear and trust students fall.

Instructors, too, may have a similar continuum that would bear investigating; however, in their case the poles may be fear of losing power at one end and trust in sharing power at the other end. Factors that may influence where instructors fall on the continuum may include prior experience with empowering students, degree of self-esteem and self-confidence, degree of concern over the amount of authority and control they need to exercise, and the general degree of trust or fear they feel toward the administration.

Another area for future research concerns determining whether a correlation exists between how actively involved students are in the learning process in class and the creation of a climate of trust as well as to what degree active and/or collaborative learning activities foster a climate of trust. In particular, it could be fruitful to study instructors who employ Classroom Assessment Techniques to solicit regular comments from students about how they are doing at promoting student learning. Furthermore, one could study the impact those

instructors have on promoting a climate of trust and whether these actions translate into more candid and sincere student ratings. Such research might use a qualitative design aimed at probing what students from both groups think about their involvement or lack of involvement in the classes. Related to this topic is conducting a study to determine if a relationship exists between instructors' pedagogical skills and practices and the climate they create in the classroom.

In addition to how actively students are involved is what impact the climate in general has. Studies could investigate whether a correlation exists between the reason for conducting ratings and the climate that knowledge of the reason creates. The question is whether knowing that the purpose of ratings is for improving teaching and learning will promote a climate of trust and, therefore, more candor and sincerity as opposed to knowing the purpose is for tenure and rehiring, thus promoting a climate of fear and lack of candor and sincerity.

One could construct a quantitative survey to determine student attitudes towards SRTs. A quantitative survey would likely capture a wider spectrum of students, including those who have no desire to participate in an interview. Furthermore, researchers might want to focus on investigating the norms that students apply when completing ratings and the impact varying norms have on the reliability and validity of SRTs. This phenomenon may also suggest that students would benefit from some training in how to undertake the rating process. In the context of this study, the examples of Greg's internal norm and Susan's external norm provide evidence that students bring preconceptions about the general norms they will apply in completing ratings as well as reflect that they probably think about their ratings rather than rating instructors based merely on their mood.

The concept of power relationships that exist between students and faculty, between faculty and administrators, and between students and administrators and the effect that power has on creating both a climate and a culture is worth more investigation. Connected with this concept is an investigation of ways in which institutions promote nutritive and integrative power.

Finally, the participants in my study came from a college in which the primary purpose for student ratings is for tenure and rehiring. A study of student attitudes at an institution in which students know that the purpose of ratings is to help instructors improve their teaching should yield different results, particularly in regard to any factors related to fear. At an institution such as Alverno College, for example, in which assessment is embedded within the college culture, I would expect students to respond quite differently about their candor and sincerity. Additionally, another group of students worth studying are those at institutions in which students control the rating process and use ratings for the purpose of “grading” instructors for student publications to assist students in selecting their professors.

Recommendations Based on the Findings

If, as this study concludes, power plays a central role in determining the climate in the classroom, not to mention the culture of the college, both faculty members and administrators might do more to exercise their power consciously in ways that will promote a trusting classroom environment. Both groups, but particularly instructors, hold the keys to fear and trust, and by enacting the following suggestions, it appears that they will be far more likely to produce trust:

1. Communicate - Explain the rating process so that students understand what it means

and what role they play in the process. At the institutional level, the administration should help students understand the process, the purpose of ratings, and the role student ratings play in the overall process. At the course level, instructors should explain how they have used student comments to improve their teaching, thus, to an extent, doing as Miles suggested when he mentioned exposing their vulnerability. In other words, instructors should apply May's notion of integrative power. By citing how past comments have helped effect improvements, instructors are showing students that they recognize that they are not perfect, that they are willing to listen and act on student suggestions, and that they want to include students in the improvement process. The message students are likely to receive is that these instructors value students—that they care. The emphasis needs to be on the value of ratings for providing constructive comments to help instructors improve teaching and learning rather than emphasis being on using ratings for the primary purpose of tenure or rehiring. Improving students' understanding of the process means improving the communication process between the administration and the students as well as between the instructor and the students.

2. Dialogue with students – Related to communication about the rating process is encouraging both instructors and administrators to engage more in true dialoguing with students. Doing so encourages critical thinking and an environment in which instructors listen to students, not just lecture to them. By engaging students in discussion and permitting the expression of critical viewpoints, the instructor has a better change of creating a climate of trust in which students understand that the instructor values the freedom of people to express their views and thus encourages openness.

3. Optimize the conditions for completing the ratings – Ensure that the ratings are

anonymous and conduct final (summative) course ratings at the end of the semester, actions that Braskamp, Brandenburg, and Ory (1984) also recommend. Additionally, announce the administration date for the SRT so that students can consider their comments prior to attending that class and to ensure better attendance on the day the students complete it. Finally, if feasible, institutions might adopt an online means of conducting SRTs, enabling students to complete their ratings on their own time from a computer terminal on campus or even from home.

4. Help students develop norms – Provide students with a clear norm for rating instruction to help prevent students applying their personal norms to try to obtain a more consistent rating. Also, consider training students in the rating process. Or perhaps the rating instruments themselves should contain questions that help obtain a sense of the criteria the students are using when they rate instructors. Although the questions on any given SRT provide aspects of instruction for students to rate, such as organization, content, enthusiasm, accessibility, etc., a student's response can vary depending on personal norms. Furthermore, participants indicated having uncertainty about the difference among the choices on the rating scale.

5. Make an effort to act on student comments – Provide students with a sense of how student ratings and their comments will improve instruction. Instructors need to clarify the purpose of student ratings for them personally and how they can benefit from the ratings. They can enhance their chances of receiving candid and sincere remarks through various methods. Instructors could relate how student comments from previous classes have helped improve their teaching. Better yet, instructors could elicit student comments early in the semester and act on the suggestions they can, letting the students know that they are making

changes based on student suggestions, thus demonstrating their sincerity in wanting to obtain student comments to help improve teaching and learning.

6. Be human...just not too human – Instructors make mistakes as all of us do. When something has not worked or is not working in a class, when an instructor makes an error, when an instructor may not know the answer to a question and is willing to admit it—such acts on the part of instructors makes them “vulnerable,” to use Miles’s word. By displaying a certain degree of humanity, instructors can generate a climate of trust, especially if instructors seek and act on student suggestions for improving the class. Being human, however, does not excuse a faculty member from being prepared and for putting forth a visible effort to teach the class. Instructors who are continually late to class, who use outdated materials, who are not timely in marking and returning student work, who admit to being unprepared, and so forth can signal to students that they really do not care about the students. These are the instructors who convey the sense that they are “just there for a paycheck.” Therefore, it appears that it would be to instructors’ advantage to admit being human as long as they are also demonstrating that they are making an effort to be the best that they can be.

7. Care about helping students learn – Develop a caring attitude that translates into treating students equitably, fairly, and with respect. The findings of this study illustrate that when instructors display a caring attitude toward students, students care in return. Such caring can translate into more trust and more candid and sincere ratings.

8. Develop the knowledge and pedagogical skills and practices that will promote student learning and critical thinking – Although participants did not provide that many examples of positive experiences with instructors, those who do, provide examples of instructors who promote active learning and use some form of Classroom Assessment

Techniques in their classes. Promoting student learning and critical thinking through active learning techniques is a means of empowering students. Again, the result is likely to be a climate of trust.

9. Reputations can change over time – Just because instructors elicit and act on student comments does not mean that they can overcome a prior reputation. A culture does not change overnight, and neither will instructors' reputations. For instructors who realize that they are the authoritarian and controlling type, those who may well be creating a climate of fear but who also want to change, they can. However, reputations may take some time to overcome. Through consistently empowering students, instructors are likely to change their reputations over time.

Final Thoughts

Greenwald and Gillmore (1998) raised the question, "Would education work better in the absence of student ratings?" Although the findings of this study show that enough doubt exists about the overall trustworthiness of ratings to support the skeptics who caution against relying too heavily on the results for the purposes of promotion and tenure, I conclude that they are likely trustworthy enough to provide some general guidance to instructors and institutions when the following conditions exist: when everyone concerned (students, instructors, and administrators) understand and promote the principle that the purpose of SRTs is to improve teaching and learning; when the process of seeking student comments begins early in the course and occurs on a regular and informal basis so as to instill a feeling that assessment and improvement are part of the classroom environment so that students learn to trust the process and how instructors will use the results; and when students feel as if instructors (and possibly also the administration) value their comments and will enact

positive changes to improve teaching based on the comments they receive. Within an environment that instills a sense of trust in students and empowers them as partners in the process of improving teaching and learning, education will work better because of student ratings. Of course, submitting to this type of scrutiny and comment takes courage on the part of the instructors as they must overcome their own fears, particularly a fear of losing their power. Instructors must be willing to listen and to avoid letting their egos stand in the way of engaging in honest reflection and appropriate change on the basis of what they hear. However, for those instructors who will relinquish their power and control, who are willing to face the possibility that perhaps they can improve with student assistance and that students can help “teach” them in this improvement process, and who take steps to become better at helping students learn, education has a much greater chance of building a culture of trust and succeeding in improving teaching and learning.

Epilogue

The following events illustrate the issue of power and the potential that exists for using power to empower students instead of asserting authority and control. I close with this story because it is a real life situation that demonstrates the concept of power at work.

Recently, an English instructor at my college approached me to ask my advice about dealing with a male student whom she described as having passive-aggressive behavior and who had become disruptive in class as well as disrespectful to her. Compounding the problem was the fact that she is African-American while he is white, so she was unsure whether his behavior might have also been racially motivated. Sitting in the middle row of seats near the front door, this student would sometimes be working on other course work during class, one day doing math, another his science. The instructor, thinking she could

employ less direct tactics to cause the student to stop, tried walking over to the space where he sat and stood there while students either discussed or asked questions about the material they had read. Although the student would stop momentarily, he began the same behavior the following week. In week four of the class, the class began engaging in a conversation about Zora Neale Hurston's "Sweat," discussing the question of race, gender, and history and how these themes shaped the characters in the short story. The student's behavior became even more blatantly disrespectful.

On Tuesday, the students were commenting Hurston's use of figurative language and how the image of a whip, which served to introduce the relationship between a husband and wife, had historical and racial implications about oppression. The instructor described what happened next:

I informed the students that the iconography of a whip could not be considered without reflecting on it as a tool of oppression against enslaved Africans. The student looked at me with a furrowed brow, squinted, and pulled a newspaper up from his desk, flipped it up in front of me, creating a popping sound as he began to read. A female student sitting next to him looked at me with perplexed eyes as if to ask was I not going to reprimand him, say something to him about what was clearly a disrespectful attitude toward not only me at this point but also the student who was speaking at the time and, therefore, to the class itself. I became so angry that it was difficult to hear how the students were addressing this concept as a reflection of the characters and the narrative itself.

Most instructors would perceive such behavior as a challenge to their authority, just as this instructor did, and it is easy to imagine a variety of responses instructors might take. The dilemma for the instructor in question was that her philosophy is to try to establish trust among her students. As she put it, "I believe that real teaching and not instructing comes from creating spaces that allow the student to expand and broaden the self, not only intellectually but socially and spiritually, so that choices and not reactive behavior can be

made.” Yet faced with such a direct challenge by this student through what she called “his act of defiance, disregard, disruption, and, most importantly, disrespect of my role as authority,” she felt that she needed to react but chose not to do it at that moment in front of the class.

Leaving the class fuming, the instructor characterized her feelings this way: “I had lost myself and gained my ego. I felt I had to exact a lesson that he would never forget.” Her plan was to come to the next class, direct a question to the misbehaving student, and as he answered the question, she would take a newspaper off her desk, flip it open, ignore what he had to say, thus “devaluing and denying his presence in front of his peers.” Then her plan was to call him out into the hallway, door open, to confront him about his problem with her so that the class would hear. One can only image what disaster would have resulted had she followed through with this action.

I did not know this was what she had been plotting when she talked with me. We both happened to be in the mailroom around 8:00 a.m., well before her class later that morning. Because we had both talked about teaching on prior occasions, she decided to ask my opinion about how to handle a situation like this. My initial response, hearing her description of the student, his behavior, and the topic of discussion and considering the fact the she is African-American, was that perhaps she should consider his gender and ethnicity and how perhaps his background might be causing conflicts within that caused him to act out. (I may work on a Ph.D. in psychology next, or perhaps play a psychologist on television.) I related that his reaction might be associated with his upbringing. Having just recently arrived at my core category for this study, it suddenly occurred to me that this was a power struggle between the instructor and the student, and I related to her my general concepts. So I asked her whether it

was more important to rid herself of this student, which would resolve the conflict between them, if it also meant losing the trust she had established with the other students by asserting her authority so flagrantly. According to my working hypothesis, by asserting herself, even though many of the students may have supported this behavior with this particular student, she would create a climate of fear, causing other students in the class to become cautious about what they would say.

My suggestion to her was to try to find a way to include him in the process and draw him into the discussion. The instructor said that by my suggesting this option, it “brought me back to my real value as an instructor, modeling growth through trust.” But she still was unsure about how she would connect with the student. All she knew was that she “had to be different, bigger, and better than what my ego demanded.”

As she walked into the learning center, she said that, as luck would have it, the disruptive student was sitting there working on his laptop computer. Seeing this as an opportunity to talk with the student outside of class in a neutral setting, the instructor explained that she had seen him doing his homework for other classes on several occasions and what she expected of students in class. He explained that he was behind in his other classes, which is why he was doing his homework in her class, and when she related how she found this behavior distracting, disruptive, and disrespectful, he responded, “I never intended to disrespect you, and I apologize for that.” She accepted his apology and told the student that the times that he did participate in class, his comments were engaging, interesting, and thought provoking, and she encouraged him to continue contributing his thoughts to the class. He told her that he had read Hurston’s story before and had “exhausted” all the ways to consider the material, relating how he thought that the comments from the other students in

class were hasty generalizations but that his philosophy is that if he cannot say anything good, it would be better to say nothing.

Now that the instructor had a better perspective of what was happening with this student, she recognized the opportunity, so she said, “Is that all that is left for you to do—criticize? Could you pose questions to get students to think about their ideas differently? I just want to leave that open to you.” The student did not respond to this suggestion verbally, but he did say that he would do his homework outside of her class from then on.

When she entered class, she decided to change the entire tone by telling the students that the short story was difficult because readers need to consider a time period and an issue that our society would rather forget, overlook, and deny but which continues to impact people today—slavery. But rather than discuss the issue of slavery or that period in history, she asked the class how the topic was affecting them and how it made them feel. Taking this approach led to a discussion about why the instructor had chosen the stories that she had, and she related how her love of literature and ideas had helped to shape her and that she had thought these works might do the same for them. The students then began making connections between the time period Hurston was writing about and today. The confrontational student chimed in, saying that Hurston showed the impact of slavery on the community and the couple in it to make the point that as individuals we should not be slaves to history. As the instructor now describes it, “He changed that day. So did I. Even though he doodles in class now and then, what he is not doing is disrupting the entire room. He is, however, engaging with fellow students and is thoughtful in speaking and writing.”

References

Abrami, P.C. (1985). Dimensions of effective college instruction. *Review of Higher*

Education, 8, 211-228.

- Abrami, P.C. (1989a). How should we use student ratings to evaluate teaching? *Research in Higher Education*, 30, 221-227.
- Abrami, P.C. (1989b). SEEQing the truth about student ratings of instruction. *Educational Researcher*, 18, 43-45.
- Abrami, P.C., Dickens, W.J., Perry, R.P., & Leventhal, L. (1980). Do teacher standards for assigning grades affect student evaluations of instruction? *Journal of Educational Psychology*, 72, 107-118.
- Abrami, P.C., Leventhal, L., and Perry, R.P. (1982). Educational seduction. *Review of Educational Research*, 52, 446-464.
- Aleamoni, L.M. (1981). Student ratings of instruction. In S. Millman (Ed.), *Handbook of teacher evaluation* (pp. 110-145). Beverly Hills, CA: Sage.
- Aleamoni, L.M. (1987). Student rating myths versus research facts. *Journal of Personnel Evaluation in Education*, 1, 111-119.
- Aubrecht, J.D. (1981). Reliability, validity, and generalizability of student ratings of instruction. Idea Paper No. 6, Center for Faculty Evaluation and Development. (ERIC Document Reproduction Service No. ED 213 296).
- Baker, G.A. (1992). Community college climate: The signature of a movement. In Baker, G.A. and Associates (Eds.) *Cultural leadership: Inside America's community colleges* (pp. 17-43). Washington, D.C.: Community College Press.
- Baker, P., & Copp, M. (1997). Gender matters most: The interaction of gendered expectations, feminist course content, and pregnancy in students' course evaluations. *Teaching Sociology*, 25, 29-43.
- Balzer, W.K., & Sulsky, L.M. (1992). Halo and performance appraisal research: A critical examination. *Journal of Applied Psychology*, 77(6), 975-985.
- Bauer, H.H. (1997). In response.... *Change*, 29, 26.
- Becker, B.E., & Cardy, R.L. (1986). Influence of halo error on appraisal effectiveness: A conceptual and empirical reconsideration. *Journal of Applied Psychology*, 71, 662-671.
- Bellah, R.N. (1999). Freedom, coercion, and authority. *Academe*, January-February, 16-21.
- Bennett, S.K. (1982). Student perceptions and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. *Journal of*

Educational Psychology, 74, 170-179.

- Bolman, L.G. & Deal, T.E. (1995). *Leading with soul: An uncommon journey of spirit*. San Francisco: Jossey-Bass.
- Braskamp, L.A., Brandenburg, D.C., & Ory, S.C. (1984). *Evaluating teaching effectiveness: A practical guide*. Beverly Hills, CA: Sage.
- Braskamp, L.A., & Ory, S.C. (1994). *Assessing faculty work: Enhancing individual and institutional performance*. San Francisco, CA: Jossey-Bass Publishers.
- Braskamp, L.A., Ory, J.C., & Pieper, D.M. (1981). Student written comments: Dimensions of instructional quality. *Journal of Educational Psychology*, 73, 65-70.
- Burdsal, C.A., & Bardo, J.W. (1986). Measuring student's perceptions of teaching: Dimensions of evaluation. *Educational and Psychological Measurement*, 56, 63-79.
- Cashin, W.E., & Downey, R.G. (1992). Using global student rating items for summative evaluation. *Journal of Educational Psychology*, 84(4), 563-572.
- Cashin, W.E., Downey, R.G., & Sixbury, G.R. (1994). Global and specific ratings of teaching effectiveness and their relation to course objectives: A reply to Marsh (1994). *Journal of Educational Psychology*, 86, 649-657.
- Centra, J.A., & Creech, F.R. (1976). *The relationship between student, teacher, and course characteristics and student ratings of teacher effectiveness*. Princeton, NJ: Educational Testing Service.
- Chacko, T.I. (1983). Student ratings of instruction: A function of grading standards. *Educational Research Quarterly*, 8, 19-25.
- Cohen, P.A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings.
- Cohen, P.A. (1983). Comment on a selective review of the validity of student ratings of teaching. *Journal of Higher Education*, 54(4), 448-458.
- Cohen, P.A. (1986). An updated and expanded meta-analysis of multisection student rating validity studies. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED 270 471).
- Cohen, S.R. (1997). The mismeasure of the college professor. *College Student Journal*, 31, 293-300.

- Coll, R. (1986). Power, powerlessness and empowerment. *Religious Education*, 81, 412-423.
- Costin, F., Greenough, W.T., & Menges, R.J. (1971). Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research*, 41, 511-536.
- Creeley, R. (1962). *For love: Poems 1950-1960*. New York: Scribners.
- Creswell, J.W. (1994). *Research design: Qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
- Creswell, J.W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.
- Cruse, D.B. (1987). Student evaluations and the university professor: *Caveat professor*. *Higher Education*, 16, 723-737.
- d'Apollonia, S., and Abrami, P.C. (1997). Navigating student ratings of instruction. *American Psychologist*, 52, 1198-1208.
- Delucchi, M. (2000). Don't worry, be happy: Instructor likability [sic], student perceptions of learning, and teacher ratings in upper-level sociology courses. *Teaching Sociology*, 28(3), 220-231.
- Delucchi, M., and Smith, W.L. (1997). A postmodern explanation of student consumerism in higher education. *Teaching Sociology*, 25, 322-327.
- Dowell, D.A., & Neal, J.A. (1982). A selective review of the validity of student ratings of teaching. *Journal of Higher Education*, 53, 51-62.
- Doyle, K.O., & Whitely, S.E. (1974). Student ratings as criteria for effective teaching. *American Educational Research Journal*, 11(3), 259-274.
- Dwinell, P.L., & Higbee, J.L. (1993). Students' perceptions of the value of teaching evaluations. *Perceptual and Motor Skills*, 76(3, Pt. 1), 995-1000.
- El-Hassan, K.A. (1995). Students' ratings of instruction: Generalizability of findings. *Studies in Educational Evaluation*, 21(4), 411-429.
- Feldman, K.A. (1976a). Grades and college students' evaluation of their courses and teachers. *Research in Higher Education*, 4, 69-111.
- Feldman, K.A. (1976b). The superior college teacher from the student's review. *Research in Higher Education*, 5, 243-288.

- Feldman, K.A. (1978). Course characteristics and college students' ratings of their teachers and courses: What we know and what we do not. *Research in Higher Education*, 9, 199-242.
- Feldman, K.A. (1979). The significance of circumstances for college students' ratings of their teachers and courses. *Research in Higher Education*, 10, 149-172.
- Feldman, K.A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education*, 18(1), 3-124.
- Feldman, K.A. (1984). Class size and college students' evaluations of teachers and courses: A closer look. *Research in Higher Education*, 21, 45-116.
- 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.
- Feldman, K.A. (1988). Effective college teaching from the students' and faculty's view: Matched or mismatched priorities? *Research in Higher Education*, 28(4), 291-344.
- Feldman, K.A. (1990). An afterward for "The Association Between Student Ratings of Specific Instructional Dimensions and Student Achievement: Refining and Extending the Synthesis of Data From Multisection Validity Studies." *Research in Higher Education*, 31, 315-318.
- Feldman, K.A. (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.
- Feldman, K.A. (1997). Identifying exemplary teachers and teaching: Evidence from student ratings. In R.P. Perry & J.C. Smart (Eds.) *Effective teaching in higher education: Research and practice* (pp. 368-395). New York: Agathon Press.
- Fisicaro, S.A., & Vance, R.J. (1994). Comments on the measurement of halo. *Educational and Psychological Measurement*, 54(2), 366-371.
- Fiske, S.T., & Neuberg, S.L. (1990). A continuum of impression formation, from category-based to individuating process: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, 23, 1-74.
- Flacks, R., & Scott, L.T. (1998, November 27). Among affluent students, a culture of disengagement. *The Chronicle of Higher Education*, A48.
- Frankhouser, W.M. (1984). The effects of different oral directions as to disposition of results on student ratings of college instruction. *Research in Higher Education*, 20,

367-374.

- Freedman, R.D., & Stumpf, S.A. (1978). Student evaluation of courses and faculty based on a perceived learning criterion: Scale construction, validation, and comparison of results. *Applied Psychological Measurement*, 2, 189-202.
- Freedman, R.D., Stumpf, S.A., & Aguanno, J.C. (1979). Validity of the course-faculty instrument (CFI): Intrinsic and extrinsic variables. *Educational and Psychological Measurement*, 39, 153-158.
- Frey, P.W., Leonard, D.W., & Beatly, W.W. (1975). Student ratings of instruction: Validation research. *American Educational Research Journal*, 12, 327-336.
- Gigliotti, R.J., & Buchtel, F.S. (1990). Attributional bias and course evaluations. *Journal of Educational Psychology*, 82, 341-351.
- Gillett-Karam, R., & Pena, E. (1992). Instructional leadership: Building a culture of excellence in the teaching-learning community. In Baker, G.A. and Associates (Eds.) *Cultural leadership: Inside America's community colleges* (pp. 97-111). Washington, D.C.: Community College Press.
- Gillmore, G.M., & Greenwald, A.G. (1999). Using statistical adjustment to reduce biases in student ratings. *American Psychologist*, 54(7), 518-519.
- Gillmore, G.M., Kane, M.T., & Naccarato, R.W. (1978). The generalizability of student ratings of instruction: Estimates of teacher and course components. *Journal of Educational Measurement*, 15, 1-13.
- Glazer, B.G. (1998). *Doing grounded theory: Issues and discussions*. Mill Valley, CA: Sociology Press.
- Greenwald, A.G. (1997). Validity concerns and usefulness of student ratings of course instruction. *American Psychologist*, 52, 1182-1186.
- Greenwald, A.G., and Gillmore, G.M. (1997a). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52, 1209-1217.
- Greenwald, A.G., & Gillmore, G.M. (1997b). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology*, 89(4), 743-751.
- Greenwald, A.G., & Gillmore, G.M. (1998). How useful are student ratings? Reactions to comments on the current issues section. *American Psychologist*, 53(11), 1228-1229.

- Griffin, B.W., & Pool, H. (1998). Monitoring and improving instructional practices (and are student evaluations valid?). *Journal of Research and Development in Education*, 32(1), 1-8.
- Haladyna, T., & Hess, R.K. (1994). The detection and correction of bias in student ratings of instruction. *Research in Higher Education*, 35, 669-687.
- Holmes, D.S. (1972). Effects of grades and disconfirmed grade expectancies on students' evaluations of their instructor. *Journal of Educational Psychology*, 63, 130-133.
- hooks, bell. (1994). *Teaching to transgress: Education as the practice of freedom*. New York: Routledge.
- 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.
- Howard, G.S., & Maxwell, S.E. (1980). Correlation between student satisfaction and grades: A case of mistaken causation? *Journal of Educational Psychology*, 72, 810-820.
- Howard, G.S., & Maxwell, S.E. (1982). Do grades contaminate student evaluations of instruction? *Research in Higher Education*, 16, 175-188.
- Jackson, D.L., Teal, C.R., Raines, S.J., Nansel, T.R., Force, R.C., & Burdsal, C.A. (1999). The dimensions of students' perceptions of teaching effectiveness. *Educational and Psychological Measurement*, 59(4), 580-596.
- Jeffreys, M.R., Massoni, M., O'Donnell, M., & Smoldaka, I. (1997). Student evaluation of courses: Determining the reliability and validity of three survey instruments. *Journal of Nursing Education*, 36, 397-400.
- Judd, C.M., Drake, R.A., Downing, J.W., & Krosnick, J.A. (1991). Some dynamic properties of attitude structures: Context-induced response facilitation and polarization. *Journal of Personality and Social Psychology*, 60, 193-202.
- Kishor, N. (1995). The effect of implicit theories on raters' inference in performance judgment: Consequences for the validity of student ratings of instruction. *Research in Higher Education*, 36, 177-195.
- Kulik, J.A., & McKeachie, W.J. (1975). The evaluation of teachers in higher education. *Review of Research in Education*, 3, 210-240.
- Lance, C.E., LaPointe, J.A., & Stewart, A.M. (1994). A test of the context dependency of three causal models of halo rater error. *Journal of Applied Psychology*, 79(3), 332-340.

- Long, G.L., and Lake, E.S. (1996). A precondition for ethical teaching: Clarity about role and inequality. *Teaching Sociology*, 24, 111-116.
- Marsh, H.W. (1980). The influence of student, course and instructor characteristics on evaluation of university teaching. *American Education Research Journal*, 19, 485-497.
- Marsh, H.W. (1982a). Factors affecting students' evaluations of the same course taught by the same instructor on different occasions. *American Educational Research Journal*, 19(4), 485-497.
- Marsh, H.W. (1982b). Validity of students' evaluations of college teaching: A multitrait-multimethod analysis. *Journal of Educational Psychology*, 74, 264-279.
- Marsh, H.W. (1983). Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics. *Journal of Educational Psychology*, 75, 150-166.
- 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. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, 253-388.
- Marsh, H.W. (1991). A multidimensional perspective on students evaluations of teaching effectiveness: Reply to Abrami and d'Apollonia. *Journal of Educational Psychology*, 83(3), 416-421.
- Marsh, H.W., & Hocevar, D. (1991). The multidimensional perspective on students evaluations of teaching effectiveness: The generality of factor structures across academic discipline, instructor level and course level. *Teaching and Teacher Education*, 7(1), 9-18.
- Marsh, H.W., & Overall, J.U. (1981). The relative influence of course level, course type, and instructor on students' evaluations of college teaching. *American Educational Research Journal*, 18, 103-112.
- Marsh, H.W., and Roche, L.A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52, 1187-1197.
- Marsh, H.W., and Roche, L.A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology*, 92(1), 202-228.

- Marsh, H.W., & Ware, J.E. (1982). Effects of expressiveness, content coverage, and incentive on multidimensional student rating scales: New interpretations of the Dr. Fox effect. *Journal of Educational Psychology*, 74, 126-134.
- Martin, E. (1984). Power and authority in the classroom. *Signs*, 9, 482-492.
- Martinson, D.L. (2000). Student evaluations of teaching and their short term validity. *Journalism and Mass Communication Educator*, 54(4), 77-82.
- Mateo, M.A., & Fernandez, J. (1996). Incidence of class size on the evaluation of university teaching quality. *Educational and Psychological Measurement*, 56(5), 771-778.
- May, R. (1972). *Power and innocence: A search for the sources of violence*. New York: W.W. Norton.
- McKeachie, W.J. (1997). Student ratings: The validity of use. *American Psychologist*, 52, 1218-1225.
- Miles, M.B., & Huberman, A.M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Miller, A.H. (1988). Student assessment of teaching in higher education. *Higher Education*, 17(1), 3-15.
- Murphy, K.R., Jako, R.A., & Anhalt, R.L. (1993). Nature and consequence of halo error: A critical analysis. *Journal of Applied Psychology*, 78, 218-225.
- Murray, H.G. (1997). Effective teaching behaviors in the college classroom. In R.P. Perry & J.C. Smart (Eds.) *Effective teaching in higher education: Research and practice* (pp. 171-204). New York: Agathon Press.
- Nasser, F., & Fresko, B. (2002). Faculty views of student evaluation of college teaching. *Assessment and Evaluation in Higher Education*, 27(2), 187-198.
- Nathan, B.R., & Lord, R.G. (1983). Cognitive categorization and dimensional schemata: A process approach to the study of halo in performance ratings. *Journal of Applied Psychology*, 68, 102-114.
- Ory, J.C. (1990). Student ratings of instruction: Ethics and practice. In M. Theall & J. Franklin (Vol. Ed.). *New directions for teaching and learning*, no. 43. *Student ratings of instruction: Issues for improving practice* (pp. 63-74). San Francisco: Jossey-Bass.
- Ory, J.C., & Braskamp, L.A. (1981). Faculty perceptions of the quality and usefulness of three types of evaluative information. *Research in Higher Education*, 15, 271-282.

- Ory, J.C., Braskamp, L.A., & Pieper, D.M. (1980). Congruency of student evaluative information collected by three methods. *Journal of Educational Psychology*, 72, 321-325.
- Palmer, L. (1998). Influence of students' global constructs of teaching effectiveness on summative evaluation. *Educational Assessment*, 5(2), 111-125.
- Perry, R.P., Abrami, P.C., & Leventhal, L. (1979). Educational seduction: The effect of instructor expressiveness and lecture content on student ratings and achievement. *Journal of Educational Psychology*, 71, 107-116.
- Phillips, S.B. (1998). Student evaluation of faculty instruction: Inflated results and student feedback. (Doctoral dissertation, University of Kansas, 1998). *Dissertation Abstracts International*, 60, 1478.
- Pike, C.K. (1998). A validation study of an instrument designed to measure teaching effectiveness. *Journal of Social Work and Education*, 34(2), 261-271.
- Powell, R.W. (1977). Grades, learning, and student evaluation of instruction. *Research in Higher Education*, 7, 193-205.
- Rubin, R. (1981). Ideal traits and terms of address for male and female college professors. *Journal of Personality and Social Psychology*, 41, 966-974.
- Salmons, S.D. (1993). The relationship between students' grades and their evaluation of instructor performance. *Applied H.R.M. Research*, 4(2), 102-114.
- Schein, E.H. (1985). *Organizational culture and leadership*. San Francisco, CA: Jossey-Bass.
- Scott, C.S. (1975). Some remarks on "student ratings: Validation." *American Educational Research Journal*, 12(4), pp. 444-447.
- Shepperd, J.C. (1997). Relevance and responsibility: A postmodern response. *Teaching Sociology*, 25, 333-335.
- Simpson, R.D. (1995). Uses and misuses of student evaluations of teaching effectiveness. *Innovative Higher Education*, 20, 3-5.
- Solomonson, A.L., & Lance, C.E. (1997). Examination of the relationship between true halo and halo error in performance ratings. *Journal of Applied Psychology*, 82(5), 665-674.
- Spencer, K.J., & Schmelkin, L.P. (2002). Student perspectives on teaching and its evaluation. *Assessment and Evaluation in Higher Education*, 27(5), 397-409.

- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage.
- Stumpf, S.A., & Freedman, R.D. (1979). Expected grade covariation with student ratings of instruction: Individual versus class effects. *Journal of Educational Psychology*, 71, 293-302.
- Trout, P.A. (1997). What the numbers mean: Providing a context for numerical student evaluations of courses. *Change*, 29, 24-30.
- Ware, J.E., Jr., & Williams, R.G. (1975). The Dr. Fox effect: A study of lecturer effectiveness and ratings of instruction. *Journal of Medical Education*, 50, 149-156.
- Waters, M., Kemp, E., & Pucci, A. (1988). High and low faculty evaluations: Descriptions by students. *Teaching of Psychology*, 15, 203-204.
- Wheatley, M.J. (1992). *Leadership and the new science: Learning about organization from an orderly universe*. San Francisco: Berrett-Koehler.
- Wheatley, M.J. & Kellner-Rogers, M. (1996). *A simpler way*. San Francisco: Berrett-Koehler.
- Wigington, H., Tollefson, N., & Rodriguez, E. (1989). Students' ratings of instructors revisited: Interactions among class and instructor variables. *Research in Higher Education*, 30(3), 331-344.
- Williams, R.G., & Ware, J.E., Jr. (1977). An extended visit with Dr. Fox: Validity of student ratings of instruction after repeated exposures to a lecturer. *American Educational Research Journal*, 14, 449-457.
- Williams, W.M., and Ceci, S.J. (1997). 'How'm I doing?' Problems with student ratings of instructors and courses. *Change*, 29, 12-23.
- Wilson, R. (1998, January 16). New research casts doubt on value of student evaluations of professors. *The Chronicle of Higher Education*, p. A12-A14.
- Worthington, A.G., & Wong, P.T.P. (1979). Effects of earned and assigned grades on student evaluations of an instructor. *Journal of Educational Psychology*, 71, 764-775.

Appendix A

Interview Protocol

Introduction:

Students being interviewed received information about the nature of the study and were able to ask questions about it if they had any. I reminded them that the purpose of this interview was not to check about specific classes or instructors; therefore, they should refrain from identifying any course or instructor. Also, I allowed the students to choose pseudonyms at that time if they preferred and reminded them that any mention of them by name in the study would be by first name or first name and last initial only.

Main Questions and Possible Probing or Follow-up Questions:

1. What do you think about or consider when you are filling out ratings?
2. Have you ever not written what you really think about a course or an instructor on a course evaluation survey, or have you ever overrated or underrated an instructor in completing an evaluation? If so, would you discuss what caused you to react as you did? [Possible follow-up questions: Would you say that this is generally true of most courses and instructors when you complete a rating, or does it vary according to the instructor? If it varies, what factor or factors cause it to vary?]
3. Under what conditions do you feel that you can't be totally open and honest when completing a rating of an instructor?
4. What could instructors and/or administrators do to encourage students to be more candid and sincere in their ratings?