

Student Personality Characteristics, Teacher Ratings,
and Student Achievement

by

Philip C. Abrami

A dissertation submitted to the Department of Psychology
in partial fulfillment of the requirements of the Faculty of
Graduate Studies for the degree of Doctor of Philosophy

University of Manitoba

Winnipeg, Manitoba

August, 1978

STUDENT PERSONALITY CHARACTERISTICS, TEACHER RATINGS,
AND STUDENT ACHIEVEMENT

BY

PHILIP C. ABRAMI

A dissertation submitted to the Faculty of Graduate Studies of
the University of Manitoba in partial fulfillment of the requirements
of the degree of

DOCTOR OF PHILOSOPHY

© 1978

Permission has been granted to the LIBRARY OF THE UNIVER-
SITY OF MANITOBA to lend or sell copies of this dissertation, to
the NATIONAL LIBRARY OF CANADA to microfilm this
dissertation and to lend or sell copies of the film, and UNIVERSITY
MICROFILMS to publish an abstract of this dissertation.

The author reserves other publication rights, and neither the
dissertation nor extensive extracts from it may be printed or other-
wise reproduced without the author's written permission.



DEDICATION

To my father, who showed me the way.

ACKNOWLEDGEMENTS

The author wishes to express his appreciation to Dr. Raymond P. Perry for his guidance and assistance throughout all phases of this research. Gratitude is also extended to Drs. H.J. Keselman, W.J. McKeachie, Les Leventhal, and Martin Samoiloff for their critical reading of this dissertation and many helpful suggestions. Connie LeBlanc did a marvelous job as typist, proof-reader, and sympathetic listener. Finally, I wish to thank my wife, Valerie, for her patience and support.

ABSTRACT

Teacher rating form (TRF) researchers may be interested in whether student personality characteristics explain variability in: (a) student descriptions about their instructor's effectiveness, or (b) the relationship between ratings and teacher-produced achievement. Such evidence may be used to judge TRF validity and generalizability. In addition, the relationship between student characteristics and achievement may be of interest.

A review of the literature suggests that student characteristics, especially those which are relevant to an educational context, appear related to ratings. Unfortunately, the results of prior studies are not conclusive. Even less can be said about the effect of student characteristics on achievement or the ratings/achievement relationship. Therefore, the present study explored the relationship between student characteristics and (a) ratings, (b) achievement, and (c) the ratings/achievement relationship. Of particular interest was the relationship between students' own (OWN) characteristics and each criterion independent of teacher (TEACHER) effects. Also of interest was the relationship between students' perceptions of instructor (PERCEPT) characteristics and each criterion independent of TEACHER characteristics, OWN characteristics, and their interaction.

The subjects were 388 male and female students from the Introductory Psychology course at the University of Manitoba. All students first completed an Adjective Check List (ACL; Gough & Heilbrun, 1967) for themselves. Students then viewed one of four colour videotaped lectures

which varied systematically in lecture content (high, low) and instructor expressiveness (high, low). Finally, students completed a TRF, quiz on the lecture, and an ACL for the videotaped instructor.

To reduce the number of student characteristics, the ACL self-ratings were factor analyzed and four ACL scales were selected for further analyses. Subsequently, the data were analyzed via stepwise multiple linear regression for each criterion measure.

The present study failed to find evidence of a meaningful relationship between ratings and student personality characteristics independent of teacher characteristics. Neither the set of four OWN characteristics, nor the set of TEACHER by OWN interactions explained a significant amount of rating variance. However, the set of PERCEPT characteristics accounted for at least 10% of the variance in ratings.

As might be expected, TEACHER characteristics explained a large percentage (30.8%) of the variance in student achievement. While several other sets of characteristics were significantly related to achievement they each accounted for less than 3% of the achievement variance. Otherwise, only the set of TEACHER by PERCEPT interactions related meaningfully to student achievement.

There was no evidence of a meaningful relationship between student personality characteristics and the ratings/achievement criterion. There was no significant OWN main effects or a TEACHER by OWN interaction. However, the set of PERCEPT characteristics was meaningfully related to the ratings/achievement criterion. This data suggests that variability in ratings and achievement can be explained by PERCEPT characteristics. Also, variability in ratings independent of achievement (or vice-versa)

can be explained by PERCEPT characteristics.

The results of this study need to be replicated in a non-laboratory setting. Several sections of a multisection course, taught by different instructors and employing a common final exam would be needed to confirm the present findings under more natural classroom conditions.

TABLE OF CONTENTS

	Page
INTRODUCTION	1
Issues in TRF Research.	2
TRF Validity Models.	2
TRF Validation Designs	7
Unit of Analysis	12
TRF Generalizability	16
Student Personality Characteristics	24
One or Two Student Characteristics	29
Several Student Characteristics	36
The Student-Teacher Match.	44
Implications of Prior Research.	55
The Student Opinion Construct.	55
The Ratings/Achievement Relationship	60
Person Perception Theory and Aptitude- Treatment Interactions.	60
Research Questions.	63
METHOD	66
Subjects.	66
Materials	66
Personality Measure.	66
Videotaped Lectures.	73
Dependent Measures	74
Procedure	75
Data Analysis	75

(Table of Contents - continued)

	Page
RESULTS	80
Factor Analysis.	80
Ratings	82
Achievement.	87
The Ratings/Achievement Relationship	89
DISCUSSION.	94
The Student Opinion Construct.	94
Student Achievement.	97
The Ratings/Achievement Relationship	98
Cautions and Suggestions for Further Research.	100
REFERENCE NOTES	102
REFERENCES.	103
APPENDIX	110

LIST OF TABLES

	Page
TABLE	
1 Simulated TRF Generalizability Data	19
2 Description of the ACL Scales	67
3 Orthogonal Varimax Rotated Factor Matrix.	81
4 Means and Standard Deviations of Predictor and Criterion Measures Used in the MRAs	83
5 Summary of Regression of TRFs and Achievement	85
6 Forward (Stepwise) MRA to Find Best Subset of Predictors.	88
7 Summary of Repeated Measures MRAs	90
8 Forward (Stepwise) Repeated Measures MRAs to Find Best Subset of Predictors	93

INTRODUCTION

There are a variety of methods for evaluating college teaching including peer, alumni, and self-ratings of instructors. One technique frequently employed and the subject of much research is teacher rating forms (TRFs) in which students at or near the end of a course usually complete a set of multiple choice questions concerning the teaching characteristics of the instructor and sometimes aspects of the course such as textbooks, examination formats, etc..

The purpose of this dissertation is to explore the relationship between college student personality characteristics and student ratings of teacher effectiveness. Many TRF researchers have investigated this relationship because they believe it provides evidence of TRF validity. Unfortunately, the interpretation of prior research is not at all clear. For example, two recent reviews of the literature on student personality characteristics-TRF rating relationships (Feldman, 1977; Follman, 1975) are not in agreement regarding the implications of existent studies. Such an inconsistency highlights the need for further study in this area.

The introduction is divided into three major sections: In the first section, four TRF research issues are discussed: (a) TRF validity models, (b) validation designs, (c) units of analysis, and (d) TRF generalizability. An understanding of these issues is important when interpreting TRF studies. The second major section includes a review of the appropriate literature. The review considers separately studies of one or two student characteristics, several characteristics, and the student-teacher match. Person perception theory is also discussed as

one way to explain the relationship between student characteristics and ratings. The final section identifies several research questions.

Issues in TRF Research

Some TRF critics cite papers which claim a correlation between student characteristics or course grades and TRF scores (e.g., Follman, 1975; Bausell & Magoon, 1971) as evidence that TRF scores are not "valid" measures of teaching ability. TRF proponents may cite the same articles and draw different conclusions or reference studies which seem to show opposite results. The purpose of this section is to demonstrate how inconsistencies in previous research may be due to: (a) the TRF validity model employed, (b) the TRF validation design used, (c) the unit of analysis employed, or (d) the TRF generalization procedure followed.

TRF Validity Models

Teacher rating form validation studies are conducted to determine what may properly be inferred from TRF scores. Such studies provide evidence of the usefulness of TRFs. A TRF validity model should describe the necessary conditions for validation research. It may, for example, dictate the relevant independent and dependent variables, unit of analysis, and research design for validation studies.

The first step in constructing a model for TRF validity is to decide on the goals of evaluation. Deciding on these goals answers the question, "What should the TRF measure?" There are many different goals against which a TRF may be evaluated. For example, (a) Do the TRF

items elicit a representative sample of student descriptions of their instructors? (b) Does the TRF predict teacher effectiveness as measured by some criterion? (c) Does the TRF predict teacher-produced differences in students on some criterion measure?

Selecting a goal of teacher evaluation can be a complex matter. A researcher may make a decision based on a variety of factors including theoretical, empirical, or practical considerations. For example, the researcher may choose a goal based on some theory of teacher effectiveness. Here, the goal of evaluation might be a TRF which measures the dimensions of effective teaching outlined by the theory. Or, the researcher may wish to develop a TRF which measures effective teacher behaviours as identified by prior research. Finally, the researcher may be guided by practical considerations such as the ease of TRF administration and scoring, faculty and student attitudes towards evaluation, and so on.

After selecting the particular goal(s) of teacher evaluation, the second step in constructing a TRF validity model is to outline the requirements of validation research. These research requirements are derived from the evaluation goal. Since evaluation goals may differ, it should be obvious that not all TRF research findings are relevant to each goal. Stating the research requirements answers the question, "What is evidence for TRF validity?" Consider the following examples of TRF validity models, goals, and research requirements.

For TRF researchers interested in representative student descriptions as the only goal of teacher evaluation it must be shown that the TRF items can elicit a representative sample of the performance domain of

interest. That is, it would be important to insure that the TRF items elicit a representative sample of all dimensions of student descriptions about instructors that are of interest to the researcher. Thus, the relationship between student ratings and student learning would not be of interest. For example, Hildebrand and Wilson (1970) surveyed 338 students for their descriptions of the characteristics of effective teachers. From the student responses a 91-item pool was generated. Factor analyses of the responses of 1015 students to the items about their instructor yielded five dimensions, each represented by several items. The Hildebrand and Wilson TRF is valid to the extent that these procedures resulted in a set of items which can elicit a representative sample of student descriptions about teachers. Such a validity model resembles the content validity model as described in the American Psychological Association (1974) test standards manual.

If, however, one were concerned with predicting teacher effectiveness as measured by student learning, the relationship between ratings and learning would be critical while the representativeness of the responses elicited by the TRF items would be unimportant. As Leventhal (1975) notes, many studies which investigated the relationship of TRF scores to some validity criterion employ the same core design. Students are non-randomly assigned to multiple sections of the same course. Towards the end of the course, students rate their instructor on a TRF and are measured on a validity criterion, such as a common final exam. Finally, a single TRF mean and a single exam mean are computed for each section, and a correlation coefficient is computed across sections between the TRF means and exam means. The correlation coefficient is

then used as evidence for TRF validity. This validity model resembles the criterion-related validity model (American Psychological Association, 1974).

A third validity model requires that TRF scores correlate with teacher characteristics that theoretically affect a validity criterion such as student achievement. Unlike the previous model, this model may, depending on the theory, require that TRF ratings predict variation in the validity criterion that is caused by teachers. Thus, finding a TRF-learning correlation would be insufficient evidence for the adequacy of the TRF unless differences in student achievement could be attributed to teachers. Such a model resembles the construct validity model (American Psychological Association, 1974).

Other TRF validity models are possible. For example, a TRF researcher might be interested in showing that representative student descriptions predict teacher effectiveness as measured by some criterion. Other TRF researchers may be interested in showing that representative student descriptions are not influenced by characteristics presumed to be unrelated to teacher effectiveness. Here, researchers are concerned with how student opinions of teachers are formed. That is, if it can be shown that student or teacher characteristics (e.g., age, sex, or "personality"), or other variables (e.g., expected grade in course), presumed to be unrelated to teacher effectiveness, affect the opinions that students form about teachers, then the validity of TRF scores may be questioned. This may occur even if the TRF scores form a representative sample of descriptions about effective teachers. Crittenden and Norr (1975) have described this approach as being within the convergent-

discriminant validation tradition (Campbell & Fiske, 1959). Crittenden and Norr suggest that when using this approach, a valid TRF should discriminate against (i.e., not correlate with), "biasing factors" not theoretically related to teaching effectiveness. Campbell and Fiske emphasize the importance of discriminant validation, suggesting that it may be equally important to find a low correlation between a test and certain unrelated measures as it is to find a high correlation between the test and the criterion.

As already noted, some validity models do not require that TRF scores correlate with a criterion such as student achievement for the TRF to be valid. If the TRF is valid and if student ratings and student achievement are uncorrelated, Whitely (Note 1) correctly suggests that rating and achievement measures may be used as independent sources of information about teaching. Unfortunately, the use of a TRF which is unrelated to some aspect of student learning would probably unsettle most college instructors and administrators. Nevertheless, validity models which do not require a ratings/achievement relationship appear to be used by some TRF researchers.

Leventhal (1975) notes that the goal of most TRF researchers is not merely to demonstrate that TRF scores correlate with some validity criterion but (due to implicit theories about how effective teachers affect students) to show that TRF scores correlate with teacher-produced differences in the criterion measure. Note that this research goal does not require that TRF items elicit a representative sample of student descriptions. Here, researchers are interested in teacher-to-teacher variation in ratings. However, an additional goal may be a TRF which is

unaffected by certain student characteristics. Here, researchers may be interested in whether ratings predict teacher-produced achievement and the characteristics which produce student-to-student variation in ratings.

In this section, some of the different models of teacher evaluation were discussed. Of particular importance was the idea that not all goals of teacher evaluation are the same. Since the goals are not the same, the methods for establishing TRF validity will differ. Thus, what is appropriate validity evidence for one model of teacher evaluation may be inappropriate for another.

TRF Validation Designs

The different validity models have generated a variety of validation designs. The following section describes weak and strong validation designs for three validity models. The evaluation goals underlying these validity models include: (a) a TRF which predicts teacher-produced achievement, (b) a TRF which elicits representative samples of student descriptions of their instructors, and (c) a TRF which elicits student opinions that have been affected by only certain teacher characteristics and not student or situational characteristics.

As already noted, the TRF/achievement correlation based on course section means (where students are not randomly assigned to sections) is typically used as evidence of TRF validity. While such a design may provide satisfactory evidence of TRF validity for some researchers, Leventhal (1975) suggests that traditional TRF validation methodology is inappropriate for researchers interested in a TRF which predicts variation in the validity criterion caused by teachers. This is because the

outcome of conventional correlational studies fails to support uniquely the notion that the TRF correlates with teacher-produced achievement. As is the case with most correlational evidence, rival interpretations are possible which either reverse the direction of the causal connection between teacher and achievement or allow for third variable explanations.

Rival interpretations become less tenable if administrative assignment of students to sections produces student equivalence from section to section. Leventhal, Abrami, Perry, and Breen (1975) examined the section selection process of students in multi-section courses at the University of Manitoba. Student responses to a questionnaire survey revealed significant section-to-section differences on three classes of variables: (1) student demographic characteristics, (2) importance of students' reasons for section selection, and (3) students' pre-enrollment information about instructor. The evidence fails to support the use of traditional TRF methodology because student nonequivalence from section to section allows for rival interpretations of the TRF/achievement correlation.

One rival interpretation, namely that student characteristics produce the correlation between TRFs and achievement, was examined in follow-up studies by Leventhal, Abrami, and Perry (1976, Note 2). Students from Introductory Psychology who had responded to the section selection questionnaire completed a TRF near the end of the course. Results indicated that: (a) students varied across sections in the importance of teacher ability/reputation to their section selection, and (b) students who selected a section because of their instructor's favorable teaching reputation rated their instructor more favorably and received

higher course grades than students not using this criterion for section selection. In such a situation, a correlation between TRF scores and achievement scores may occur across sections means without teacher affecting achievement and be mistakenly used as evidence of TRF validity.

According to Leventhal (1975), stronger methodologies than the traditional correlational design exist for showing whether TRFs predict teachers who promote achievement. A stronger design is an experimental design that requires random assignment of students to teachers. Support for TRF validity can then be shown by computing a single correlation coefficient across teachers between TRF means and examination means. Random assignment of students to sections makes it unlikely that student characteristics will affect the correlation between TRF scores and student achievement. When random assignment of students cannot be used, repeated measurements correlational (panel) designs may be employed. Unlike the conventional correlation and experimental designs, panel designs require that TRF and achievement measures be gathered at least twice during a course. Analysis of panel design data allow either the teacher affects student achievement or the student achievement affects teaching hypothesis to be rejected. In addition, third variable interpretations can be tested using techniques such as partial correlation analysis. While not as strong as the randomized subjects design, panel designs have greater internal validity than conventional correlational designs which are subject to both third variable and reverse direction explanations.

A second validity model has as its goal TRFs whose items elicit a representative sample of student descriptions of their instructors.

For this TRF validity model, the researcher must: (a) define the content universe (e.g., students' opinions about teachers) and (b) demonstrate that the TRF elicits a representative sample of the content universe. In defining the universe, the TRF researcher must defend the appropriateness and thoroughness of the definition. For example, the researcher may wish to define the "student opinions" in terms of only some of the opinions students have about teachers (e.g., lecture organization) and not others (e.g., teacher rapport). In sampling from the content universe, the TRF researcher must also defend the sampling technique. For example, do nonanonymous, written opinions about teachers accurately reflect students' private opinions? In this case, the appropriate validation design would compare students who were randomly assigned to different opinion sampling techniques (e.g., anonymous vs. nonanonymous) where teacher and situational effects were held constant. Any teacher effect might be controlled by selecting students from only one class or, if multiple classes (i.e., teachers) are used, a stratified random assignment of students to the opinion sampling techniques might be used where an equal number of students from each class are randomly assigned to each experimental group (i.e., a teacher by anonymity factorial).

A third validity model is concerned with how student opinions are formed. For this model, characteristics presumed to be related to teacher effectiveness (probably only certain teacher characteristics) must be shown to affect TRF scores. Other characteristics (probably many student characteristics, situational characteristics, and some teacher characteristics), not related to teacher effectiveness, must be

shown not to affect TRF scores. To investigate whether teacher characteristics affect TRF scores, student and situational characteristics must be equivalent across teacher conditions. This may be accomplished by randomly assigning students to teacher conditions (i.e., classes) and holding situational characteristics constant. To investigate student effects on TRF scores, situational and teacher effects can be controlled by either employing a single class (consisting of students from all levels of the student characteristic) or, if multiple classes are used, employing a stratified random assignment of students to classes and holding situational effects constant (i.e., a teacher by student factorial).

This later design assumes that a heterogeneous class (made up of students with different levels of the student characteristic will yield similar results to homogeneous classes (made up of students with the same level of the student characteristic). If situational effects and teacher effects are constant across classes (e.g., the instructor does not modify his teaching behaviour from class to class because in one class he has bright students and in another dull students) differences between the two designs will occur only if there are different student effects. For example, a class of only dull students may award low TRF ratings because no one appears to be learning. However, the same group of dull students might award high TRF ratings when placed in a class with bright students because they no longer attribute their poor performance to the teacher. When such outcomes are likely, the choice of homogeneous versus heterogeneous classes should depend on the situation to be generalized to.

To this point, appropriate and inappropriate validation designs for three different TRF validity models have been described. The design of

a validation study is related to the model of interest. But even when a validation study is properly designed, the interpretation of the findings can be misleading if the results are not properly analyzed.

Unit of Analysis

Following the selection of a TRF validity model and appropriate validation designs, another choicepoint for TRF researchers is the correct unit of analysis. Recently, Leventhal, Abrami, and Perry (Note 2) have addressed this issue. They attempted to show: (a) how the same data can be analyzed in several ways depending on the unit of analysis employed, and (b) that the interpretation of the results may markedly differ as a function of the unit of analysis chosen.

Their comments were addressed particularly to researchers who study multisection courses to determine the correlation between ratings and teacher-produced achievement. However, their arguments concerning the unit of analysis can be extended to other evaluation goals. In the following sections, the appropriate units of analysis are described for the following goals of evaluation: (a) a TRF which predicts teacher-produced achievement, (b) a TRF which correlates with achievement, (c) a TRF which elicits representative student descriptions, and (d) a TRF which elicits students' opinions that have been affected by only certain teacher characteristics.

For multisection courses, Leventhal et al. described three different units of analysis used by researchers: classmeans, students within a class, and pooled students. The "classmean" is the mean score for the students in a section. "Students within class" are the individual scores for the students in one section only. In many studies where students

within class are the units of analysis, analyses are computed separately for each section and the results averaged. "Pooled students" are the individual scores for the students from all sections of a multi-section course.

For researchers interested in knowing whether TRFs predict teacher-produced achievement, Leventhal et al. suggest that the classmean is the appropriate unit of analysis. When students are randomly assigned and situational factors are held constant, variation in achievement and rating classmeans is attributable to teacher effects and random variation. Other units of analysis, such as students within a class or students pooled from all classes are inappropriate. Using students within a class as the units of analysis may result in variation in ratings and achievement which are attributable to student differences, not teacher differences, since the effect of the teacher is held constant. Similarly, the use of students pooled across classes as the units of analysis may result in achievement variation which is attributable to both students and teachers.

Seldom in TRF validation research are students randomly assigned to classes. Unfortunately, when students are not randomly assigned to classes, variation in achievement and rating classmeans is no longer attributable solely to teachers and random variation. Class-to-class differences in students may also contribute to achievement and rating variation. In such instances, some TRF researchers have attempted to remove the effect of class-to-class differences in student ability from the achievement classmeans by using ability pretest scores to residualize achievement scores. Even if the ability pretest is valid, and student ability is the only factor on which sections differ--a tenuous assumption-- a problem may arise if the student effect, as measured by the

ability pretest, on achievement classmeans is correlated with the teacher effect on achievement classmeans. Using ability pretest classmeans to residualize achievement classmeans not only will remove the effect of class-to-class differences in student ability, but also may remove any teacher effects on achievement which happen to covary with student ability. Since there is no way in this design to determine the true effect of teachers on achievement classmeans, one can only say that the true correlation of TRF scores with teacher-produced achievement scores lies somewhere within the range of the TRF classmean-residualized achievement classmean correlation (i.e., the "semi-partial" or "part correlation") and the TRF classmean-unresidualized achievement classmean correlation.¹

For researchers interested in whether TRF scores correlate with achievement, either classmeans, students within class, or students pooled across classes are acceptable as the unit of analysis. Since it is not important to determine what produced the achievement scores that

¹Other TRF researchers, interested in how student opinions are formed as well as predicting teacher-produced achievement, may want to use the ability pretest to residualize TRF scores as well as achievement scores. When both TRF scores and achievement scores are controlled for student ability, the correlation of TRF scores and achievement scores is referred to as the "partial correlation." For these researchers it may be desirable to remove any possible source of TRF variation due to student ability so that variation in TRF classmeans can be attributable to teacher effects only. For example, imagine that in a nonrandomized design, class-to-class differences in student ability were entirely responsible for variation in TRF classmeans, and that variation in achievement classmeans was attributable to teachers only. Here, TRF scores may predict teacher-produced achievement, not because student ratings were affected by teacher performance, but because class-to-class differences in ratings due to students happened to correlate with class-to-class differences in achievement due to teachers. Unfortunately, the partial correlation suffers interpretive ambiguities more complex than the semi-partial correlation. That is, not only will the partial correlation remove any teacher effects on achievement classmeans which covary with ability classmeans, teacher effects on TRF classmeans which covary with ability classmeans will also be removed.

correlated with the TRF, any unit of analysis may be employed. Of course, even in the same study, computing these three correlations can yield different correlation coefficients.

For researchers interested in whether TRF scores elicit representative student descriptions of their instructors, the appropriate unit of analysis is typically the mean of student scores within each student opinion condition (i.e., the subclass or cell mean). For example, a researcher might be interested in the effect of anonymity on TRF scores in a particular class. Here, one compares the subclass mean of scores on anonymously completed TRFs to the subclass mean of scores on non-anonymously completed TRFs, taking into account within subclass variance. This is statistically computed with Student's t-test between means.

The subclass mean or classmean is typically the appropriate unit of analysis for researchers interested in how student opinions are formed. When it is important to determine the effects of student characteristics on TRF scores and if heterogeneous classes are employed, the subclass mean is the appropriate unit of analysis. However, for researchers interested in the effect of teacher, situational, or student characteristics using homogeneous classes, the classmean is the appropriate unit of analysis.

As a general procedure for choosing the appropriate unit of analysis one should ask several questions: (a) what is the relationship of interest? (i.e., what are the relevant independent and dependent variables and how should they be reflected?) (b) What are the sources of variability in the dependent measure(s)? (c) Which of these sources are of interest, which are to be controlled? For example, suppose that the

effect of student sex on TRF scores was of interest. Here, a student variable, which is unrelated to teaching effectiveness, is examined for its effect on the formation of student opinions about their teacher. Sources of rating variability might include student sex and other student characteristics and, if there is more than one class, teacher and situational characteristics. Since we are interested only in the effect of student sex on ratings, we would choose a unit of analysis where differences in ratings could be attributed to student sex and random variation.

TRF Generalizability¹

The generalizability of student ratings is of concern to some TRF researchers. The purpose of the following discussion is to show how certain studies may or may not provide evidence of generalizability for the following evaluating goals: (a) a TRF which predicts teacher-produced achievement, (b) a TRF which elicits representative student descriptions, and (c) a TRF which elicits student opinions that have been affected by only certain teacher characteristics.

What is TRF generalizability? Studies of TRF generalizability are concerned with determining the conditions under which student ratings are valid. For example, will the TRF be equally useful for chemistry courses as psychology courses? Generalizability is never directly measured but inferred from the research evidence; it is judged as satisfactory or otherwise. Studies may evaluate generalizability across three sets of variables: student characteristics, teacher characteristics, and situational characteristics. Researchers have attempted to

¹Parts of this section are based on a paper by Abrami, Leventhal, and Perry (Note 3).

determine if TRF scores are influenced by characteristics of student raters because student characteristics such as age, sex, and ability vary within and between classes. Since teacher and situational characteristics vary between classes, it is also important to study TRF generalizability across these sets of factors.

Typically, TRF generalizability studies merely examine the relationship between TRF scores and student, teacher, or situational characteristics. For example, Doyle and Whitely (1974) report data to suggest that student ratings are predominantly unrelated to student characteristics. Unfortunately, such findings may or may not be evidence of generalizability depending, among other things, on the model of TRF validity employed.

For researchers interested in whether ratings predict teacher-produced differences in the criterion measure, a suitable generalizability study would investigate the relationship between student characteristics and the correlation (or some other measure of association) between TRF scores and teacher-produced student achievement (or some other criterion). Here, a properly designed generalizability study is a validity study repeated for each level of the student characteristic to which one wants to generalize. Given a properly designed study, judgments of TRF generalizability depend further on the researcher's criteria for generalizability. For example, McKeachie (Note 4) has suggested that a TRF is generalizable as long as the ratings/achievement correlations are positive across the levels of the student characteristic. Other researchers might claim TRF generalizability only if the correlations are positive and not substantially different across the levels of the student characteristic.

Consider the simulated data presented in Table 1. Cases 1 and 2 represent data comparable to Doyle and Whitely's (1974). That is, the student characteristic does not correlate with TRF scores. But the TRF would probably be judged as generalizable in Case 1 only. In Case 1, the correlations, computed for each level of the student characteristic, between mean TRF scores and mean achievement scores are not different. In Case 2, the student characteristic affects student achievement in such a way that the correlations between TRF scores and achievement vary substantially. This result is not unlikely with such student characteristics as intellectual ability or motivation to learn. For example, classes of bright students may perform well with little regard to the efforts of their instructor but the classes may also award different TRF scores depending on certain instructor characteristics. Alternately, classes of dull students may perform poorly with little regard to the efforts of their instructors, yet give out very different teacher ratings.

Doyle and Whitely concluded that TRF scores are unaffected by student characteristics and are therefore generalizable. The simulated data for Cases 1 and 2 suggest that this inference may be incorrect if the ratings/criterion correlation varies due to student characteristics. Doty (1967) demonstrated that various student characteristics correlate with student achievement. Hence, the relationship between TRF scores and achievement scores may vary due to student characteristics.

Cases 3 and 4 represent simulated data for a student characteristic that does correlate with TRF scores. Nevertheless, the TRF is generalizable in Case 3 because the correlations between student ratings and achievement scores are unaffected by different levels of the student

TABLE 1

Simulated TRF Generalizability Data^a

LEVEL OF STUDENT CHARACTERISTIC	TEACHER	STUDENT CHARACTERISTIC DOES NOT AFFECT TRF			STUDENT CHARACTERISTIC AFFECTS TRF		
		MEAN TRF SCORES	MEAN ACHIEVEMENT SCORES	CORRELATION	MEAN TRF SCORES	MEAN ACHIEVEMENT SCORES	CORRELATION
HI	A	1.0	1.0	$r = 1.0$	1.1	1.1	$r = 1.0$
	B	2.0	2.0		1.2	1.2	
	C	3.0	3.0		1.3	1.3	
	D	4.0	4.0		1.4	1.4	
	E	5.0	5.0		1.5	1.5	
MED	A	1.0	1.0	$r = 1.0$	3.1	3.1	$r = 1.0$
	B	2.0	2.0		3.2	3.2	
	C	3.0	3.0		3.3	3.3	
	D	4.0	4.0		3.4	3.4	
	E	5.0	5.0		3.5	3.5	
LO	A	1.0	1.0	$r = 1.0$	5.1	5.1	$r = 1.0$
	B	2.0	2.0		5.2	5.2	
	C	3.0	3.0		5.3	5.3	
	D	4.0	4.0		5.4	5.4	
	E	5.0	5.0		5.5	5.5	
CASE 1. TRF IS GENERALIZABLE				CASE 3. TRF IS GENERALIZABLE			

Table 1 continued...

LEVEL OF STUDENT CHARACTERISTIC	TEACHER	STUDENT CHARACTERISTIC DOES NOT AFFECT TRF			STUDENT CHARACTERISTIC AFFECTS TRF			
		MEAN TRF SCORES	MEAN ACHIEVEMENT SCORES	CORRELATION	MEAN TRF SCORES	MEAN ACHIEVEMENT SCORES	CORRELATION	
HI	A	1.0	1.1	$r = 1.0$	1.1	1.0	$r = 1.0$	
	B	2.0	1.2		1.2	2.0		
	C	3.0	1.3		1.3	3.0		
	D	4.0	1.4		1.4	4.0		
	E	5.0	1.5		1.5	5.0		
MED	A	1.0	3.5	$r = -1.0$	3.5	1.0	$r = -1.0$	
	B	2.0	3.4		3.4	2.0		
	C	3.0	3.3		3.3	3.0		
	D	4.0	3.2		3.2	4.0		
	E	5.0	3.1		3.1	5.0		
LO	A	1.0	5.3	$r = .2$	5.3	1.0	$r = .2$	
	B	2.0	5.4		5.4	2.0		
	C	3.0	5.1		5.1	3.0		
	D	4.0	5.2		5.2	4.0		
	E	5.0	5.5		5.5	5.0		
				CASE 2. TRF IS NOT GENERALIZABLE				
				CASE 4. TRF IS NOT GENERALIZABLE				

^aTable 1 shows mean TRF and mean achievement scores for students who are high, medium, and low on a student characteristic and who are assigned to different teachers. Scores in the table apply for either of two kinds of student assignment: (1) heterogeneous classes where each teacher (A through E) receives high, medium, and low students, or (2) homogeneous classes where each teacher receives high students, then teaches another class of medium students, then another of low students. The choice of homogeneous versus heterogeneous classes for investigating TRF generalizability may depend upon the specific field situation of interest.

characteristic. For example, academic level of students was found to correlate with TRF scores (Gage, 1961). Nevertheless, one might suspect a TRF to be generalizable since academic level of students should also correlate with student achievement.

When TRF scores are generalizable but student characteristics affect the scores (Case 3), the ratings can be made comparable from teacher to teacher by "correcting" them for the student characteristics. It would be misleading to compare the ratings of two professors whose classes were different with regard to important student characteristics without adjusting the ratings. Once the validity of a TRF has been established for each level of a student characteristic, TRF scores can be corrected for student characteristics in much the same way as IQ scores on intelligence tests are corrected for age of testee.

The preceding discussion has illustrated how to determine whether a TRF which is designed to measure teacher-produced student achievement is generalizable across students. In addition to student characteristics, concern is often expressed that student evaluations should not be influenced by the personality characteristics of the instructor. This issue is, in part, a definitional one concerning the exclusion of personality characteristics from the construct "effective instruction." Researchers who argue that an instructor's personality should not affect the ratings he/she receives imply that personality style is not a dimension of good teaching. Consider the TRFs based on factor analytic studies (e.g., Isaacson, McKeachie, Milholland, Lin, Hofeller, Baerwaldt, and Zinn, 1964) which do include personality dimensions (or their behavioural correlates) such as "rapport" or "interaction." Here, the

implicit assumption is made that teacher personality may be an important part of effective instruction.

Studies which attempt to show a relationship between teacher personality and student ratings typically do not include a student achievement measure (e.g., Sherman & Blackburn, 1975). Hence, such studies fail to provide evidence of whether TRFs predict teacher-produced student achievement. To demonstrate that a TRF does not generalize, it must be shown that either: (1) differences in teacher personality cause student ratings to vary and these differences are unrelated to student achievement, or (2) differences in teacher personality cause student achievement to vary and these differences are unrelated to student ratings. In other words, teacher personality must affect the relationship between TRFs and achievement for TRF generalizability to be affected.

In addition, researchers may also be interested in TRF generalizability across situational variables (e.g., class time, class size, or different universities). The design and analysis of such studies is similar to that presented for generalizing across student and teacher characteristics. Evidence of whether TRFs predict teacher-produced student situational factors is gained by examining the effects of setting variables on the ratings/criterion correlation.

In conclusion, for TRF researchers interested in the relationship of TRF scores to teacher-produced student achievement, appropriate generalizability studies should include both a TRF and an achievement measure. Since in the past, these researchers have failed to study the generalizability of the ratings/criterion correlation, the generalizability of TRFs is still an open question.

But researchers may have other evaluation goals and, therefore, may require different procedures for determining the generalizability of student ratings. For researchers whose goal is a TRF which elicits representative student descriptions, a criterion measure is not employed so a ratings/criterion correlation is never computed. For these researchers showing that the representativeness of the responses elicited by TRF scores is unaffected by the characteristics one wants to generalize across, is sufficient evidence of TRF generalizability. For example, a TRF which is generalizable across student sex must elicit equally representative responses from male and female students. If the TRF elicits a fair sample of male student descriptions but not female descriptions, it would not be generalizable across student sex. This might occur when male opinions of effective instruction differ from female opinions.

For researchers interested in showing that student descriptions are not influenced by characteristics presumed to be unrelated to teacher effectiveness, finding that TRF scores are unaffected by the characteristics one wants to generalize across is sufficient evidence of TRF generalizability. For example, Doyle and Whitely's (1974) research provides good evidence that student descriptions of effective teaching are not affected by certain student characteristics.

In the first part of the Introduction, four issues of TRF research were discussed: (a) TRF validity models, (b) TRF validation designs, (c) unit of analysis, and (d) TRF generalizability. First, it was noted that not all goals of teacher evaluation are the same and, therefore, the methods for establishing TRF validity may differ. What is appropriate

evidence for one goal of teacher evaluation may be inappropriate for another goal. Second, it was suggested that the design of a validation study is related to the validity model of interest. Examples of strong and weak designs for several different models were given. Third, the correct units of analysis for a variety of teacher evaluation goals were presented. The different units of analysis include pooled students, students within class, and classmeans. Finally, appropriate procedures for exploring the generalizability of student ratings were discussed for different evaluation goals.

The major thrust of the prior sections was to outline some of the theoretical and methodological choicepoints of TRF researchers and to show how their decisions might substantially affect the results and interpretation of TRF research in general. Hopefully, this discussion will serve as an aid to understanding some of the apparent inconsistencies in the results and interpretations of prior research on student evaluation of instruction. In the second part of the Introduction, an attempt will be made to demonstrate how these issues relate to the research on student personality characteristics and teacher evaluations.

Student Personality Characteristics

For this paper student personality characteristics refer to "the personality traits, interests, preferences, opinions, attitudes, and values of students" (Feldman, 1977, p. 242). Excluded from this review are studies which deal with student demographic characteristics (e.g., age or sex), student abilities, motivations, grade expectations, and certain student initial impressions (e.g., good-bad teacher), or pre-

course impressions about a particular course or instructor. The interested reader should consult Feldman's (1977) excellent review of these latter studies. Studies which investigated only characteristics that are measured following the commencement of a course and which are likely to have been affected by the teacher or the course, are not considered. For example, a study by Granzin and Painter (1973) which investigated: (a) student attitudes toward a course at the end of the semester, and (b) change in attitude toward a course from the beginning to the end of the semester, would be excluded from consideration. Also, only studies which employed some form of student evaluation as an outcome measure will be considered. Hall's (1970) use of a perceived learning instrument (with such items as "How satisfied do you feel about this course?") would satisfy this criterion, whereas Painter and Granzin's (1972) use of grade expectations (which may not measure a student's impression of the instructor at all) would not. Finally, parts of studies irrelevant to the effects of student characteristics on rating validity will be excluded.

TRF researchers do not agree on the extent to which student characteristics affect TRF validity. Elsewhere, an attempt was made to delineate general issues in TRF validation research which may contribute to this inconsistency. That these issues are relevant to student characteristics is exemplified in Feldman's (1977) review. With regard to the question of which student characteristics should be considered biasing elements in ratings, Feldman states:

The matter is more complex really, being contingent on whether ratings are objective, or subjective, or a mixture of both. If, on the one hand, ratings are meant or claimed to be objective, then ideally

none of the attributes of the students nor any of their class experiences should be related to ratings. Any that do are biasing results. Clearly students' anticipated grades in the course, interest and motivation brought to the course, certain predispositions, and the like, should not be related to objective ratings of the teacher or course. Neither, for that matter, should motivation, interest, and learning induced by the teacher. Even if some students were more inspired by the teacher and learned more from him or her (or thought they did), these experiences should not affect neutral descriptions and assessments of the teacher's degree of preparation and organization of the course, knowledge of the subject matter, or any of the specific areas in which students are asked to rate their teachers. Indeed, it is arguably the case that these class experiences also should not be related to overall ratings of the general "effectiveness" of the teacher, if these global ratings are meant to be objective assessments.

If, on the other hand, ratings are the subjective assessment of the teacher--either theoretically or in practice--then teacher-inspired motivation and teacher-induced learning would be expected to be associated with students' evaluation of the teacher's overall "effectiveness" as well as some of the more specific areas of the teacher's performance. Considered more generally, other of the student's characteristics and experiences might also be expected to correlate with the student's evaluation of the teacher, given the general theory and research on the personal factors that affect individuals' perceptions and evaluation of the qualities and behaviors of others (1977, pp. 252-253).

Feldman's (1977) comments about the objectivity-subjectivity of ratings relate to a prior discussion concerning TRF validity models. There, it was pointed out that researchers interested in the effect of student characteristics on ratings may be concerned with the validity of the student opinion construct. These researchers want to know how student opinions are formed--whether, for example, certain student characteristics affect students' judgments of instructor effectiveness. There may be other researchers, however, interested in the correlation of ratings with teacher-produced achievement, who might not be interested in how student opinions are formed or whether ratings are "objective" or "subjective." For these researchers, a TRF must correlate with teacher-produced achievement to be valid. Thus, if student characteristics do

not affect the ratings/achievement relationship, whether student characteristics affect ratings alone is of less concern (see section entitled "TRF Generalizability").

Rather than address both types of TRF evaluation goals, Feldman's review is directed at researchers primarily interested in the validity of the student opinion construct. This is clear from examining his comments on the unit of analysis problem:

Moreover, since interest lies in the correlates of the variability among individual students in the same classes, only studies in which the individual student is the unit of analysis are included and not studies in which the class or course itself is the exclusive unit of analysis. This procedure is important, for the two types of studies essentially ask and answer different questions (cf. Menzel, 1950). It is generally hazardous to draw inferences about the direction and strength of relationships at the group level of analysis (i.e., "ecological analysis") from the direction and strength of relationships at the individual level of analysis (i.e., "individual analysis"). (Feldman, 1977, p. 235)

In addition, Feldman contrasts his remarks about the objectivity-subjectivity of ratings when the individual student is the unit of analysis to studies employing classmeans as the unit of analysis: "The argument is different at the group level of analysis. Given certain controls, positive associations between average teacher or course ratings and the average indicators of student achievement in the content of the course are generally expected (and usually found)." (1977, p. 260)

Feldman's (1977) remarks on the relationship of student characteristics to TRF ratings are based on three kinds of studies: (a) those which include one or two student characteristics, (b) those which include a variety of student characteristics, and (c) those which investigate the match between students and teacher. For the first two kinds of studies, Feldman concluded that: "Studies of one or two of these sorts of

personality and related characteristics have not found them to be related to ratings of courses and teachers . . . By contrast studies that have used a number of dimensions and indicators of the personality, attitudes, or values of students (and usually a number of different rating items or scales) have found certain associations between the characteristics and ratings" (1977, p. 242, p. 243). He went on to suggest, however, that for some of these studies the number of significant results might be expected by chance alone.

In striking contrast to these remarks by Feldman are the conclusions of Follman (1975) who also reviewed the literature on "the specific effects of student raters' personality characteristics on their ratings of faculty teaching effectiveness." (p. 156)

The weight of the empirical evidence indicating a relationship between raters' personality characteristics and their ratings of teachers' effectiveness is overwhelming. The evidence shows that different raters have different personality characteristics, that rater personality characteristics influence substantially their ratings of instructors' teaching effectiveness, and that raters' personality characteristics influence teacher ratings differentially. The best guess regarding a maximum correlation between a univariate personality variable and a teacher rating is about 0.25. The best estimate regarding a characteristic correlation between a multivariate set of personality variables and a teacher rating is a multiple correlation of about 0.45. (p. 163)

These contrasting remarks underscore the need to carefully examine those studies which investigated the relationship of ratings to a few or many student characteristics.²

A second area of interest is the studies which provide evidence of

²Part of this discrepancy may be explained by the fact that the two reviewers do not cite the same studies, which may in turn be a function of the difficulty of obtaining unpublished papers. Consequently, the present review will cover only published material.

the effect of student characteristics on the ratings/achievement relationship. Elsewhere, it was noted that many TRF researchers are interested in whether ratings predict teacher-produced achievement. These researchers should be interested in those student characteristics which affect the ratings/achievement relationship.

The present review will be organized into three parts. In the first part, individual studies will be organized into three categories: (a) studies which included only one or two student characteristics, (b) studies which included a variety of student characteristics, and (c) studies which investigated the student-teacher match. Within each category, studies will be further subdivided into those that include a TRF only and those which employ both ratings and achievement. In one important way, the classifications are arbitrary: all the studies deal with the effect of student characteristics on TRF validity. Research on the student-teacher match is treated separately because of the seemingly different orientation toward the problem presented in these studies. The remaining studies are divided on the basis of the number of student characteristics studied, rather than the type of characteristic studied. In the second part, the results and implications of all the studies for TRF validity will be summarized. Finally, person perception theory will be employed as a possible explanation of why student characteristics might affect TRF validity. In addition, there will be a brief discussion of student characteristic effects on achievement.

One or Two Student Characteristics

Feldman (1977) observed that studies in this category have not generally found ratings to be related to student characteristics. This

category of research is further divided here into those studies that employ a TRF only, and those including both a ratings and achievement measure.

Ratings only. Maney (1959) examined the relationship between student authoritarianism, measured by the California F scale, and student ratings, measured by summing student responses to a TRF. Students were selected from the classes of eight instructors so that each instructor was rated by four high, medium, and low authoritarian students. The correlation between student authoritarianism and ratings was nonsignificant ($r = .00$), as was the student authoritarianism/ratings correlation controlling for teacher authoritarianism ($r = .02$). Maney (1959) concluded that: "degree of authoritarianism of the student is not significantly associated with favorableness of evaluation." (p. 231). It is likely that these correlations were based on students pooled from all eight classes. Therefore, variability in ratings could reflect student, teacher, and situational effects, changing the likelihood of finding a significant correlation.

Freehill (1967) studied the rated college experience of 52 students who were either high or low authoritarian, as measured by a modified version of the California F scale. Among other findings, Freehill reported that high authoritarians were more critical of instruction than low authoritarians. Freehill concluded that "the use of student ratings to compare college teachers . . . is short of respectable science" (1967, p. 19). This conclusion is unwarranted if one considers that: (a) students did not rate specific instructors but college teachers in general, and (b) teacher differences, not student differences, might

explain the findings (e.g., low authoritarian students may have collected in the classes of better instructors).

White and Wash (1966) examined the relationship between student need for social approval and TRF scores. In each of the eight classes students were divided into two groups on the basis of their scores on the Marlowe-Crowne social desirability scale. Separate factor analyses were performed on the TRF data (Veldman and Peck's Pupil Observation Survey) for each group. Following this, the 10 factor scores for the low need group were correlated with the 12 factor scores for the high need group. Fifty of 120 correlations were significant at the .05 level. White and Wash concluded that social desirability contributes to the variance in student ratings because the factor structure of the two need groups, "markedly differed." Yet they acknowledged that the significant correlations indicated that "there appeared to be much communality between the factor structures" (1966, p. 717). Thus, it is not clear from this study the extent to which need for social approval affects student ratings.

Kovacs and Kapel (1976) explored the relationship between student need for achievement (as measured by the Mehrabian Achievement scale) locus of control (as measured by the Rotter's IE scale), and TRF scores (as measured by Kapel's Instructor Evaluation Form and a semantic differential scale). Students ($n = 286$) from the classes of 16 instructors completed the personality measures and the TRFs during a regular class meeting. Pearson product-moment correlations were computed between the student characteristics and student ratings using pooled students as the unit of analysis. Significant correlations ($p < .05$) were found between the student characteristics and both the semantic differential rating

scales and student global ratings on Kapell's form (but not specific factors from this form). Unfortunately, since pooled students were the units of analysis rather than students within classes, it is difficult to determine from this study the true relationship between student characteristics and variability in ratings produced by students. In this study, variability in ratings might be caused by student, teacher, and situational effects. Thus, it is possible that a significant correlation might exist between student characteristics and ratings caused by teachers, and this evidence might mistakenly be used as evidence of TRF invalidity. On the other hand, the lack of significant relationship between student characteristics and TRF scores may be due to the additional rating variability caused by teacher and situational effects which is unrelated to student characteristics.

Ratings and achievement. Domino (1971) studied the academic performance and teacher ratings of students who had different achievement orientations (as measured by subscales of the California Psychological Inventory). Students who were high in achievement-via-conformance (AC) were assigned to one of two classes taught by the same instructor in either a "conforming" manner or "independent" manner. Students who were high in achievement-via-independence (AI) were assigned to two other classes taught by the same instructor which also differed in presentation style. There were, therefore, four classes of homogeneous students with regard to achievement orientation. The design of the study was an achievement orientation (AI, AC) by lecture style (conforming, independent) factorial. Student satisfaction was measured by two items designed to evaluate the course and the instructor. Student achievement was

measured by: (a) a multiple choice final exam, (b) essay questions rated for factual knowledge, and (c) essay questions rated for original thinking. Separate analyses on each dependent measure revealed a fair amount of consistency--there were significant style by achievement orientation interaction effects on both student rating measures and two of three achievement measures. If one is willing to assume that the teacher did not alter his teaching style within each style condition because of class-to-class differences in students (perhaps a tenuous assumption), the data suggest that: (a) student opinions about teachers may vary due to achievement orientation, and (b) achievement orientation appears not to affect the relationship between ratings and teacher-produced achievement.

The effect of student abstract-concrete personality on ratings and achievement under different methods of instruction was investigated by Tuckman and Orefice (1973). Student personality structure was determined from scores on the Interpersonal Topical Inventory - Scale IV. For each personality subgroup, 14 students were randomly assigned to one of four classes differing in method of instruction. Student ratings were based on a four item TRF, while the achievement measure covered content common to each instructional method. Results of the personality by lecture method analysis of variance of ratings revealed a significant interaction. A similar analysis of achievement data showed no significant effects. The results of this study suggest that student concreteness-abstractness affects the ratings of an instructor, and may affect the relationship between ratings and teacher-produced achievement.

The relationships between student objectivity-subjectivity (as

measured by the Blass scale), ratings and achievement were explored by Blass (1974). Within a single class midterm test scores were not significantly different for objective versus subjective students, nor were there any significant differences between the groups on the 11 TRF items. Thus, this study fails to provide evidence that objectivity-subjectivity affects ratings or the ratings-achievement relationship.

Tobias and Hanlon (1975) studied the relationship between students' needs for social approval (as measured by a short form of the Marlowe-Crowne scale) and ratings, and the student's desire to enroll in future courses taught by their instructor (which can be considered a criterion of teacher effectiveness). Data were collected from a total of 158 students enrolled in seven classes and correlations computed between social desirability scores, a six factor McKeachie TRF, and the teacher effectiveness criterion, using pooled students as the units of analysis. No significant correlations were found between the social desirability scores and the TRF factor scores, or the criterion. Contrary to White and Wash (1966), the authors concluded that: "the results of this study provide evidence for the validity of the use of student ratings" (Tobias & Hanlon, 1975, p. 407). But for researchers interested in either the effect of social desirability on student opinions or the relationship of ratings with some teacher-produced criterion, this study fails to provide convincing evidence. Pooled students are the incorrect units of analysis for either purpose.

Page and Roy (1975) and Page (1976) studied the relationship between internal-external locus of control and the correlation of TRF scores and achievement (as measured by expected grades and actual grades,

respectively). For students from three different sections of a class, correlations were computed separately for internals and externals between two rating items and grades. The results showed that the ratings/grades correlations were significantly different for externals than internals. However, this study fails to provide convincing evidence for researchers interested in the effect of locus of control on either ratings, or the rating/achievement correlation. No evidence is presented of the effect of locus of control on ratings alone and the appropriate unit of analysis--classmeans--was not used in computing the ratings/achievement correlation.

Parent, Forward, Canter, and Mohling (1975) also studied the relationship between locus of control and ratings and achievement. The design of the study was a factorial with two levels of locus of control (internal, external), two levels of class discipline (high, low), and two levels of class preference (congruent, incongruent). Subjects were randomly assigned to class discipline conditions. The same instructor was used to teach both discipline conditions which were two hours in duration. An unspecified questionnaire assessed student satisfaction and a 30-item quiz was used to measure achievement. They found: (a) a significant interaction effect of locus of control and discipline on achievement, but not ratings, and (b) a significant interaction effect of class preference and discipline on ratings, but not achievement. Thus, locus of control did not seem to affect student opinions in this study although under certain conditions class preference did. In addition, both class preference and locus of control may have affected the ratings/achievement relationship.

Several Student Characteristics

In contrast to studies of one or two characteristics, Feldman (1977) observed that studies of several student characteristics have sometimes found associations between characteristics and ratings. However, he added that some of these associations may occur by chance alone. This category of research is further divided here into studies that employ a TRF only, and studies that include both a ratings and achievement measure.

Ratings only. Yonge and Sassenrath (1968) studied the relationship between student personality characteristics and ratings for three instructors. A nine-factor TRF and a single overall teaching ability item were used to measure student ratings. In addition, students completed the Omnibus Personality Inventory Form F which yielded 14 measures of student personality characteristics related to an academic context. Within each of the three classes, correlations were computed between the ratings and the personality scores. As Feldman (1977) notes, only a small percentage of all possible correlations were significant (approximately 15%). In addition, it was generally the case that the size and direction of the correlations varied from instructor to instructor. Yonge and Sassenrath suggest that "the type of student who tends to rate one instructor high on a given factor may be the type of student who tends to rate another instructor low on the 'same' dimension" (1968, p. 51). Thus, no evidence was found in this study of a consistent relationship between ratings and student characteristics across instructors. Rather, a student by teacher interaction was related to ratings.

Rees (1969) asked 65 college students to rate 11 types of college teachers (e.g., English teachers, art teachers, etc.) on sets of 20 semantic differential scales. For each student the 220 responses were summed and then a factor analysis using the 65 total scores was computed. Rees suggests that these factor scores: "represented 'idealized individuals' or 'points of view' in the ratings of college teachers" (1969, p. 478). The scores for eight factors were correlated with a variety of student characteristics including several personality measures (i.e., Pederson's Personality Inventory, California F Scale, Negative California F Scale, Tolerance-Intolerance of Ambiguity Scale, Category Width Scale, and the Social Desirability Scale). Only 17 of 224 correlations were significant at the .05 level. Rees concluded that: "Different students conceptualize teachers in a variety of different ways depending upon the configuration of personality and background factors existing for each type of student" (1969, p. 480). But the data would seem to suggest that this is an unjustified conclusion. Ratings of teachers or, more precisely, teachers in general did not appear to be highly related to student characteristics in this study.

Grush and Costin (1975) studied the relationship between four dimensions of student personality (as measured by the Gordon Personal Profile Inventory, or the Gordon Personal Inventory) and student ratings (as measured by five items from the skill factor of the McKeachie TRF). Since the data were collected from a large number of students from many classes, correlations between personality scores and ratings were presented using the classmean as the unit of analysis. None of these correlations were significant ($p > .05$). In addition, the authors

report that the results using individual students as the unit of analysis (probably pooled students) were very similar. This study offers no positive evidence of student characteristics-ratings relationships.

Rezler (1965) explored the relationship between student needs and student ratings. Students from ten classes completed the 15 factor Edwards Personal Preference Schedule, the 10 factor Purdue Rating Scale for Instructors, and an item which measured their perception of the instructor's feeling toward them. Regression analysis using pooled students as the units of analysis were computed using student ratings as criterion variables. With the data analyzed separately for males and females, nurturance positively correlated with ratings for males, succorance negatively correlated with ratings for females, and exhibition and heterosexuality correlated positively with ratings for males, but negatively for females. As was previously noted (see comments for Kovacs & Kapel, 1976), it is difficult to determine from studies using pooled students as the units of analysis the relationship between student characteristics and ratings. In the present study, for example, results are presented which seem to suggest that needs correlate with ratings but correlate differently for males than for females. First, it seems possible that any correlations between needs and ratings using pooled students might be incorrect because variability in pooled ratings cannot be attributed to student factors alone. Second, the differences between the correlations for males and females might be due to teacher or situational factors since students were not randomly assigned to classes (e.g., females might

be heavily enrolled in English composition and males in finance).

Tetenbaum (1975) also investigated the relationship between student needs, measured by Jackson's Personality Research Form (PRF), and ratings, measured by a single overall ability item. The PRF used had previously been factor analyzed to arrive at 12 scales which represented four need factors. Brief vignettes describing a college classroom were prepared so that a hypothetical teacher's behaviour corresponded to one of the four needs. In all, 12 vignettes were created so that there were three replications of each teacher orientation. A total of 405 students completed the PRF and rated each vignette on the single TRF item. Canonical correlations were computed between the 12 PRF scales and the 12 vignettes. Three significant canonical correlations were found. Subsequent analysis of the correlations between the PRF scales and the three canonical variates revealed that, generally, student needs were related to ratings of teacher orientations congruent with those needs and not others. Thus, a specific student need was not found which could account for rating variability consistently across teacher orientations. Specific student needs predicted teacher ratings only when they were compatible with the teacher's orientation (i.e., a needs by teacher orientation interaction).

Several studies have explored the relationship between student values or attitudes and ratings. Feldhusen and Starks (1970) studied the relationship between student attitudes toward courses and

professors in general and ratings of a particular course or instructor. At the end of the semester ratings were collected for the particular instructor. The correlations between generalized attitudes and specific ratings, while significant, were all low (.16 to .34). Feldhusen and Starks concluded that the ratings of an instructor are not determined by students' general attitudes toward courses and instructors.

Crittenden and Norr (1973) were also interested in the relationship between student attitudes toward teaching and student ratings. In their study a 24-item TRF and three overall effectiveness items were administered to 1,718 students in 52 classes. Students were asked to rate each item in terms of its importance for quality teaching and then to evaluate their instructor. For each item on the TRF, students were divided into high, medium, and low groups according to how important they judged that aspect of good teaching. Then, correlations were computed for each group between the specific item ratings of the instructor and the overall effectiveness ratings. An examination of these correlations suggested that correlations between specific items and overall evaluations were positively related to the importance assigned to the particular characteristic. That is, for each characteristic the correlations were generally lowest for the low importance group, and highest for the high importance group. Crittenden and Norr suggested that when the importance of a specific factor increases, its weight in the overall evaluation must increase and, therefore, be more highly correlated with overall

effectiveness ratings. They conclude that:

These results also point to a serious drawback in using students' overall course or instructor evaluation ratings as a basis for ranking instructors with respect to teacher quality. Since these global evaluations are systematic products of the interaction of student values and teacher behavior, they hardly provide an uncontaminated account of teacher performance. (Crittenden & Norr, 1973, p. 150).

The causal connections that Crittenden and Norr used to arrive at this conclusion were that: (a) overall evaluations are the weighted average of specific evaluations, and (b) the weights of specific items are determined by their importance to the rater. As they note, however, the "lack of experimental control makes inference about the perceptual process an untidy and difficult task." (Crittenden & Norr, 1973, p. 146).

Other interpretations of their findings are possible. One alternative suggests that students concerned about particular teacher characteristics seek out particular instructors. In a study such as this, where students are not randomly assigned to classes, certain students may collect in particular classes. Since pooled students are the units of analysis, the student values affect ratings explanation could be supplanted by a teacher characteristics affect ratings explanation. Here, students may seek out particular instructors but assign ratings on the basis of the instructor's performance rather than their own values. In fact, evidence that students do seek out particular instructors and selectively register in their courses was found by Leventhal, Abrami, Perry, and Breen (1975).

Null and Walter (1972) explored the relationship between student values in general (measured by the Allport, Vernon, Lindzey Study of Values (AVL)) and student ratings (measured by the first ten items of

the Purdue Rating Scale for Instructions). Students (N = 192) from one class completed both the AVL and TRF toward the end of the semester. For each of the six dimensions of the AVL students were placed in one of two groups depending on their factor scores. The data were then analyzed for each TRF item (a total of 60 statistical tests of interest). Only four tests reached significance at the .05 level. Null and Walter concluded that most dimensions of student values do not affect student opinions about instructors.

Student academic self-concept was studied by Haslett (1976). She first administered two multi-item instruments designed to measure academic self-concept and the concept of a good teacher. Factor analyses revealed a five factor solution for self-concept and a five factor solution for good teacher. Regression analyses were then performed for each good teacher factor using self-concept scores and student sex as the predictors. The total variance accounted for in student judgments of good teachers by academic self-concept and sex, ranged from 1% to 10%. Haslett suggests that academic self-concept correlated significantly with factors used by students to evaluate good teaching. However, the amount of variance accounted for by these relationships was generally minor.

Ratings and achievement. Only one study (Doty, 1967) investigated the relationship between several student social and achievement needs, creativity, ratings, and achievement under different teaching method conditions. Achievement need was measured by a scale from the Edwards' Personal Preference Schedule. Social need was measured by the social-ability scale from the Guilford-Zimmerman Temperament Survey. Creativity was measured by Getzel and Jackson's four tests of creativity. Student



ratings were the sum of scores on ten items while achievement was measured by a 75-item quiz. After completing the student characteristics questionnaires, but before filling out the TRF and quiz, students were exposed to one of three teaching methods (lecture-discussion, small group discussion, or taped lecture) administered by the same instructor. Correlations were computed separately for males and females among all the variables for each teaching method. For the three teaching method conditions, there was no consistent finding of significant within-class correlations between student characteristics and ratings for either males or females. So, this study provided no evidence that student characteristics consistently (i.e., across student sex and teacher method) relate to the way student opinions about teachers are formed. On the other hand, it was generally found that social and achievement need scores correlated with achievement. That is, students with different needs appeared to learn different amounts. In addition, the sign of the correlations infrequently varied with student sex or teaching method.

Does this study offer evidence on the relationship between student characteristics and the correlation of ratings with teacher-produced achievement? Liberally interpreted, the results suggest that: (a) there is no clearcut relationship between student characteristics and student opinions; (b) there is a relationship between student characteristics and student achievement; and, therefore, (c) the relationship between ratings and teacher-produced achievement may depend on the way student characteristics vary from class to class. These data are similar to Case 2 in Table 1 where, for the same instructor, student

characteristics do not affect ratings but affect achievement. However, because (a) correlations between mean class ratings and achievement for different levels of student characteristics were not computed and (b) the students may not have been randomly assigned to teaching method conditions; results of the Doty study do not provide strong evidence that the ratings/achievement relationship varies due to student characteristics.

The Student-Teacher Match

A variety of researchers have been interested in the match, congruence, or similarity of students and teachers as it relates to student ratings. But the concept of student-teacher similarity can be operationally defined in a variety of ways. For example, a similarity score can be based on: (a) the difference between a student's own personality score and the instructor's own personality score (actual similarity), (b) the difference between a student's own personality score and the class average rating of the instructor's personality (another actual similarity score), or (c) the difference between a student's own score and that student's rating of the instructor's personality (perceived similarity). Differences in operational definitions of similarity scores plus certain design considerations may create interpretative problems, especially when attempting to integrate the findings of different similarity studies.

Computed within class, the perceived similarity scores/ratings correlation might identify a student characteristic which relates to ratings. That is, the manner in which students perceive themselves and their instructor either as similar or dissimilar--may relate to ratings. An alternative to correlating perceived similarity scores and

ratings is to regress TRF scores on the two variables which constitute a perceived similarity score: the student's own score (O) and the student's rating of the instructor's personality (P). The correlation of perceived similarity and ratings, or the regression of ratings on perceived similarity, will always be equal to or less than the regression of TRF scores on O and P. The regressions will be equal when either O or P are constants. As Hays (1973) notes, adding (or subtracting) a constant value to each value of a variable does not change its variance. Otherwise, the regression of ratings on perceived similarity will be less than the regression of ratings on O and P. Note that the regression of ratings on O and P equals the regression of ratings on perceived similarity and P (which also equals the regression of ratings on perceived similarity and O). Regressing ratings on O and P rather than perceived similarity may account for more variance in ratings. In addition, one variable (O or P) may account for most of the rating variance. Depending on the researcher's needs, future studies might incorporate only this variable. In contrast, studies of perceived similarity must measure both O and P.

The prior example focused on the linear relationship between perceived similarity and ratings. However, a researcher might be interested in knowing if there was any relationship (linear and nonlinear) between perceived similarity and ratings. For example, imagine that O and P were dichotomous variables with values 0 and 1. In this case, perceived similarity could take on three values (1, 0, -1) labeled, say, positively, dissimilar, similar, and negatively dissimilar. A one-way analysis of variance (ANOVA) could be computed with ratings as the dependent variable and perceived similarity as the independent variable. The

F-test for perceived similarity would have two numerator degrees of freedom (dfs). Alternately, one could compute a two-way ANOVA which tests 0, P, and 0 x P. These are all tests with one numerator df so only linear relationships are tested. However, the ratings variance accounted for by perceived similarity is either equal to that accounted for by 0 and P, or 0 and 0 x P, or P and 0 x P, or a subset of that accounted for by 0, P, and 0 x P. In other words, the variance in ratings accounted for by perceived similarity in a one-way ANOVA will always be equal to or less than the total variance explained in an 0 x P ANOVA.

Similar logic applies when 0 and P are multilevel variables. For example, imagine that both 0 and P had three levels: 1, 2, and 3. Here perceived similarity scores could be -2, -1, 0, 1, or 2. The F-test from the one-way ANOVA for perceived similarity has four numerator dfs. Alternately, one could compute a two-way ANOVA which tests 0, P, and 0 x P. The variance in ratings accounted for by perceived similarity may be equal to that accounted for by 0 and P (4 dfs), or 0 x P (4 dfs), or a subset of the variance accounted for by 0 and 0 x P (6 dfs), or P and 0 x P (dfs), or 0, P, and 0 x P (8 dfs). As in the bilevel case, the variance in ratings accounted for by perceived similarity is less than or equal to the total variance accounted for in a multilevel 0 x P ANOVA. There may be other benefits to conceptualizing and analyzing perceived similarity studies as 0 x P factorials. As already mentioned, one variable (0 or P) may account for most of the rating variance, reducing the complexity of future studies.

Computed across pooled students, the perceived similarity scores/ratings correlation might identify a student characteristic which

relates in part, to teacher-produced ratings. Thus, pooled students would not be appropriate units of analysis. Similarly, classmeans, with students randomly assigned to classes, would be inappropriate units of analysis because variability in both ratings and perceived similarity scores would not be attributable to students.

For actual similarity scores (scores based on the difference between students' own scores and the instructor's own score or the class average) the similarity scores/ratings correlation computed within a class is equivalent to the correlation between student's own characteristics and ratings. Here, the instructor's own score and the class average are constants, so subtracting either from students' own scores will leave the latter variance unchanged.

The actual similarity scores/ratings correlation may also be computed across classmeans. When students are randomly assigned to classes, the actual similarity scores/ratings correlation should not differ significantly from the instructor's own scores/ratings correlation (or the class average scores/ratings correlation). Here, differences from class-to-class in mean student personality scores should occur by chance alone. When students are not randomly assigned to classes, the actual similarity scores/ratings correlation may be significantly different from the instructor's own scores/ratings correlation (or the class average scores/ratings correlation). As previously described, designs where students are not randomly assigned to classes are generally to be avoided.

The actual similarity scores/ratings correlation may also be computed across pooled students. Imagine that students were assigned to classes of teachers with different personalities on the basis of the

students' own personalities (i.e., a student personality by teacher personality factorial). The actual similarity scores/ratings correlation provides evidence of whether the actual similarity between student and teacher relates to ratings. An alternative to correlating actual similarity scores and ratings is to regress TRF scores on the two variables which constitute an actual similarity score: the student's own score (0), and the instructor's own score (1) (or the class average score (A)). The regression of ratings on actual similarity will always be equal to or less than the regression of TRF scores on 0 and 1 (or A). The second advantage of treating actual similarity as two separate variables is doing so allows for rating variance to be explained in more traditional, and perhaps simpler, terms. For example, as separate variables one might describe the variance in ratings explained by teacher personality characteristics or student personality characteristics. On the other hand, one might describe rating variance explained by actual similarity only in terms of the difference between student and teacher personality. Here, one must consider student and teacher characteristics simultaneously.

As described below, the majority of studies have investigated the relationship between perceived or actual similarity and ratings. Only one study (Menges, 1969) investigated the effect of similarity on the ratings/achievement relationship.

Ratings only. Hall (1970) conducted a study to determine if the discrepancy between students' ideal teacher style and perceived teacher style could account for variability in students' reported learning. Students were asked to rate their ideal teacher and their own instructor

on a seven factor TRF. Students also completed a six factor perceived learning instrument, which can be considered as a second TRF. Within class correlations were computed between (a) the own instructor ratings and the perceived learning instrument, and (b) the discrepancy scores and the perceived learning instrument. It was generally found that the discrepancy scores did not predict perceived learning any better than own instructor ratings. Furthermore, it was noted that students appeared to give the ideal teacher high ratings on all dimensions with little variability. Thus students were fairly unanimous in their ratings of the ideal teacher. In other words, a student characteristic was not identified in this study which either varied substantially within a class or related to ratings.

Levinthal, Lansky, and Andrews (1971) were also concerned with the relationship between students' ideals and ratings. They had 263 students from one class rate their own instructor and their ideal instructor on a nine-item adaptation of the McKeachie TRF. Except for one item, it was found that "ideal responses do not tend to be unanimous" (Levinthal et al., 1971, p. 107). In contrast to Hall, these students varied within class in evaluating their ideal instructor. Evidence concerning the relationship between student ideals and actual teacher ratings was, unfortunately, incomplete. Data for only three items was presented and then only a moderate relationship was found between ideal and actual ratings on two of these items.

Davison (1973) studied the relationship between perceived similarity among instructor and student values and ratings. Students in two sections of the same course taught by the same instructor rated

themselves and the instructor on Bill's Index of Adjustment and Values. Then, students rated their instructor on the 19-item Teaching Effectiveness Scale. Perceived similarity was determined by comparing the student's average self-rating to the average of his/her ratings of the instructor's values. Students who perceived the instructor as superior to themselves had an average instructor evaluation which was significantly different than students who did not have this perception. Assuming no between section effects, the data support the hypothesis that perceived value similarity affects ratings.

Good and Good (1973a) explored the effects of attitude similarity on ratings of an hypothetical instructor. Students filled out a 14-item Survey of Attitudes dealing with diverse topics. Several days later, each student received an attitude survey which had been completed by an hypothetical instructor in such a way that it either expressed similarity or dissimilarity with the students' own views. Students were asked to evaluate the hypothetical instructor on the six-item Instructor Evaluation Scale. For each TRF item, the similar group was significantly different than the dissimilar group. If one assumes that the assignment of students to the similar-dissimilar attitude conditions was random, then the data suggests that actual similarity of attitudes with hypothetical instructors affects student opinions.

Levenson and LeUnes (1973) conducted a followup study to determine if the same findings as Good and Good would occur with a real instructor. They had the students in a single class complete the attitude survey and rate the instructor on the TRF used in the prior study. Attitude similarity was determined by comparing student values to the instructor's

own values. For this instructor significant differences between similar and dissimilar students were reported for three of four items. In contrast to the prior study, this study held teacher values constant. Here, differences between similar and dissimilar students are due to student attitudes only.

Good and Good (1973b) conducted a second study to examine the relationship between perceived versus actual attitude similarity and student ratings. They administered an attitude survey to 409 students in 21 classes. Students completed the survey for themselves and the instructor. Each student also rated his/her instructor on a 16-item TRF, while each instructor completed the attitude survey. Using pooled students as the units of analysis, no significant correlations were found between degree of actual similarity and course ratings. However, 13 of 16 correlations between perceived similarity and ratings were significant. Unfortunately, the results of this study for perceived similarity cannot be unambiguously interpreted due to the criticisms raised previously against the use of pooled students.

Costin and Grush (1973) explored the relationship between student ratings and: (a) students' self-described traits, (b) students' perception of their teacher traits, and (c) the discrepancy between actual and preferred teacher traits. Personality traits were measured by Gordon's Personal Profile or the Gordon Personal Inventory. Ratings were gathered on a five-factor TRF from students in the classes of more than 50 students. All data were analyzed using the classmean as the unit of analysis. Note that since students were not randomly assigned to classes and situational effects were not controlled, variability in rating

means may reflect teacher, student, and situational factors. Six of the 20 correlations between students' self-ratings on the four factor personality inventory measure and the five factor TRF were significant at the .05 level. By contrast, the students' ratings of the instructors' personality traits were more closely related to their teacher effectiveness ratings. Discrepancy scores were computed by finding the absolute difference between the students' mean preference for their teachers' trait, and the students' mean observation of their teachers' trait. Correlations were computed between the discrepancy scores for the four personality factors and only two of the four teacher effectiveness dimensions. Of these eight correlations, six were significant at the .05 level. These correlations however are difficult to interpret. First, Costin and Grush note that the variances of preferred traits were consistently smaller than the variances of observed traits. Thus, one might infer that there are not great differences between students in preferred traits and that variability in mean effectiveness ratings might be explained just as well by observed traits alone. Second, do differences from class-to-class in mean discrepancy scores reflect student, teacher, or situational characteristics? Since one cannot answer this question and since the same unanswerable question applies to TRF scores, the discrepancy scores-TRF scores correlations may be produced by any one of these sources (or some combination of sources).

Grush, Clore, and Costin (1975), using the data from another study (Grush & Costin, 1975), explored the effect of student similarity to instructor on attraction. Students from a large number of classes rated themselves and their instructors on personality inventories. Based on a

comparison of their own ratings to their instructor's mean ratings, students were classified as either similar or dissimilar on each trait. Teachers were also classified as either high, medium, or low depending on their mean scores. From each class, an equal number of similar and dissimilar students were selected. Thus, the basic design of the study was a teachers (high, medium, low) by actual similarity (similar, dissimilar) factorial. In addition, students and teachers were classified on two relevant traits and two irrelevant traits where relevancy was determined by the traits relationship to a measure of teacher skill. Thus, students and teachers were classified in four ways and four separate analyses were run (each analysis consisted of a different sample of students since not all students or teachers were used). In each case, the dependent measure was the student's rating of the teacher's personality on the trait of interest. For all traits, there was a significant teacher by personality effect. For the relevant traits only, actual similarity affected rating of the teacher's personality. Positively dissimilar students (i.e., students whose own ratings were lower than the instructor's) were more attracted to their instructor than similar students. There were no significant interactions. Thus, this study offers some support to the hypothesis that actual similarity affects ratings.

Fulcher and Anderson (1974) also investigated the relationship between actual similarity and TRF scores. They first developed a 12 factor interpersonal dissimilarity scale. Then, students from three classes were asked to rate themselves and their instructor on the scale as well as to complete two measures of overall teacher effectiveness. The findings revealed a relation between TRF scores and actual similarity across

the three classes. However, no support was found for the hypothesis that moderately dissimilar teachers would be rated as most effective.

Morstain (1977) explored the relationship between ratings and student orientations toward education, faculty orientations, and student-faculty incongruence. Both the student and faculty orientations surveys contain six almost identical factors which relate to instructional orientation. The Student Instructional Report was the TRF employed from which four factors were analyzed in the present study. Average within class correlations for nine classes between the six factor scores on the student orientation survey and the four TRF factor scores were low (ranging from .01 to .22). Disparity scores (actual similarity scores) were based on the average class student orientation on each dimension minus the instructor's self-rating. None of the correlations between disparity scores and mean class TRF scores were significant. As was the case with the Costin and Grush (1973) study these correlations are difficult to interpret. Do class-to-class differences in disparity scores reflect differences in student orientations, or faculty orientations, or both? Also, does variability in TRF scores reflect student, teacher, or situational effects?

Ratings and achievement. Only one study (Menges, 1969) investigated the relationship between student-teacher compatibility (as measured by scores on the Cognitive Structure Questionnaire) and ratings and achievement. Scores on the structure questionnaire for students in two sections of a course taught by the same instructor were compared with the instructor's own scores. Based on these actual similarity scores, three compatibility groups were formed. It was found that cognitive

compatibility related to some course and teacher evaluations whereas achievement (i.e., course grade and lecture quizzes) was not related. However, the actual similarity (compatibility) scores did not vary due to a teacher variable (there was only one instructor's self-rating). Therefore, the compatibility findings can be explained in terms of variability in students alone.

Implications of Prior Research

In a prior section, four issues in TRF research were discussed: (a) TRF validity models, (b) TRF validation designs, (c) unit of analysis, and (d) TRF generalizability. It was suggested that TRF researchers do not always have the same goals for teacher evaluation and, therefore, do not always employ the same methods for establishing TRF validity. For researchers interested in the validity of the student opinion construct, the relationship between student characteristics and ratings should be studied with students within class as the unit of analysis. For researchers interested in the relationship between ratings and teacher-produced achievement, the effect of student characteristics on the TRF validity coefficient should be examined, with the class-mean as the unit of analysis for homogeneous classes. Considering these two goals for teacher evaluation separately, the existing literature provides limited information on the effects of student characteristics on TRF validity.

The Student Opinion Construct

Most of the research on student characteristics and ratings has been concerned with the validity of the student opinion construct. In

fact, Feldman's (1977) review is concerned exclusively with these studies. An examination of the published literature reveals that research which explores the relationship between one, two, or several student characteristics and ratings is generally non-experimental (i.e., field-correlational), where students are not randomly assigned to classes, and typically does not employ the appropriate unit of analysis. Thus, interpretation of many of these studies is difficult. This literature as a whole has not identified any student characteristic or set of characteristics which the cautious researcher would say provide strong evidence of an effect on ratings. That is, a characteristic has not been found which correlates with ratings in more than one study where alternative explanations of the results are minimized. On the other hand, this statement does not warrant the conclusion that student opinions are unaffected by student characteristics. Thus, the major implication of prior research is that further research and study are necessary since strong statements about student characteristic effects cannot be made at present.

These studies, however, do suggest that a possible relationship may exist between ratings and those student characteristics which are conceptually related to the educational environment. On the other hand, there is less evidence supporting a characteristics/ratings relationship for student factors which are not as readily associated with the educational environment. Broadly speaking, this distinction suggests that factors related to a student's academic behavior (e.g., motivation to learn, interest in school, desire for good grades, etc.) may have an association with ratings while factors unrelated to academia may not.

Examples of factors related to the educational environment include: need for achievement, achievement orientation, general attitudes toward courses or teachers, and academic self-concept. Examples of factors less easily associated with the educational environment include: authoritarianism, locus of control, objectivity-subjectivity, and need for social approval. This distinction is, of course, fraught with difficulties and is offered only because it may have some heuristic value. For example, on the basis of accumulated evidence, objectivity-subjectivity may eventually be shown to be more related to factors within the educational environment than, say, academic self-concept. More importantly, this distinction does not imply that studies of student characteristics from either category will, or will not, provide sufficient evidence concerning TRF validity in the future. Rather, the distinction is one way to summarize prior research and one way to suggest a direction for future research.

A review of the studies which can be described as not readily associated with the educational environment reveals a moderately consistent picture provided one is willing to accept the evidence these studies provide as suggestive. As already noted, methodological difficulties with most of these studies make precise interpretation difficult.³ These studies provide little evidence of a relationship between student variables and ratings. Studies of student authoritarianism either suggest

³For the summary of prior research Feldman's (1977) practice will be followed, of attributing variability in ratings to students in studies which employed pooled students as the units of analysis. The reader is reminded that alternative explanations are possible.

no relationship with ratings (Maney, 1959; Rees, 1969) or are not easily interpretable (Freehill, 1967). Need for social approval or social desirability has also not been shown to affect ratings in two studies (Rees, 1969; Tobias & Hanlon, 1975) while another study is unclear (White & Wash, 1966). Locus of control findings are inconclusive: Kovac and Kapel's (1976) study suggests a relationship; Parent, Forward, Canter, and Mohling's (1975) study does not; and Page's study (Page, 1976; Page & Roy, 1975) is not readily interpretable. Blass (1974) found no significant differences in ratings between objective and subjective students. Null and Walter's (1972) research suggests no relationships between student values in general and TRF scores. Studies exploring the relationships of various personality characteristics with ratings (Doty, 1967; Grush & Costin, 1975; Rees, 1969; Yonge & Sassenrath, 1968) typically do not report large or consistent correlations. However, some evidence suggests the possibility of interaction effects of student needs and other variables (e.g., teacher style, student sex) on ratings (Doty, 1967; Rezler, 1965; Tetenbaum, 1975).

In toto, studies of student characteristics more readily associated with the educational environment provide more positive evidence of a relationship of student variables with ratings than other studies. For example, Kovac and Kapel (1976) found a relationship between need for achievement and TRF scores; Domino (1971) found that students' achievement orientation could account for some variability in ratings; and Tuckman and Orefice (1973) reported a relationship between concrete-abstract thinking and ratings. In addition, Crittenden and Norr (1973), and Feldhusen and Starks (1970) provide evidence of at least a small

relationship between students' attitudes and values towards courses or teachers and ratings. In two of three studies of student needs, which established at least an interaction effect of needs and other variables on ratings, factors which seemed related to the educational environment (e.g., intellectual striving from Tetenbaum, 1975; achievement need from Doty, 1967) were included. On the other hand, there are several studies which fail to show relationships between these relevant student characteristics and ratings. For example, Haslett (1976) found no large relationship between student academic self-concept and TRF scores, while the studies by Grush and Costin (1975), Rees (1969), and Yonge and Sassenrath (1968) do not report large or consistent correlations between relevant personality dimensions and ratings.

As a separate issue, the perceived and actual similarity between students and teachers as a source of ratings variability has received attention from a variety of researchers. Unfortunately, some previous similarity studies provide inappropriate, incomplete, or undecipherable data. Hall (1970) could not identify a student characteristic which varied substantially within a class. Levinthal, Lansky, and Andrews (1971) presented incomplete data. Good and Good (1973b) correlated perceived similarity scores and ratings using pooled students as the units of analysis. Both Costin and Grush (1973) and Morstain (1977) used classmeans as the units of analysis. Fulcher and Anderson (1974) did not report correlations, either within class or for pooled students, between perceived dissimilarity and ratings. Finally, LeUnes and LeUnes (1973) computed actual similarity scores for a single instructor.

Of the remaining studies, three examined the relationship between

actual similarity and ratings while only one studied perceived similarity. Davison (1973) found a relationship between perceived value similarity and ratings. Both Grush, Clore, and Costin (1975) and Good and Good (1973a) found a relationship between actual similarity (personality and attitude, respectively) and ratings. However, Good and Good's (1973b) second study reports no relationship between actual attitude similarity and ratings. Thus, overwhelming evidence of a relationship between either perceived or actual similarity and ratings is not presently available, suggesting further research is needed.

The Ratings/Achievement Relationship

For researchers interested in the relationship between student characteristics and the ratings/achievement relationship, only eight studies could be found which include both a TRF and a criterion measure. Of these, three studies (Page, 1976; Page & Roy, 1975; Tobias & Hanlon, 1975) present data inappropriate to conclude that a student characteristic relates to TRF validity. Of the remaining studies, two report data which suggest no relationship between student characteristics and TRF validity: Domino's (1971) investigation of achievement orientation and Blass' (1971) investigation of objectivity-subjectivity. Three studies offer some support for the claim that student characteristics relate to the ratings-criterion relationship. These studies suggest that locus of control (Parent, Forward, Canter & Mohling, 1975), creativity, social need, achievement need (Doty, 1967), and concreteness-abstractness (Tuckman & Orefice, 1973), may affect TRF validity.

Person Perception Theory and Aptitude Treatment Interactions

This portion of the paper briefly considers why student character-

istics might relate to ratings and achievement.

Why student characteristics might relate to ratings. Several authors (e.g., Crittenden & Norr, 1973; Follman, 1975; Kovacs & Kapel, 1976; Null & Walter, 1972; Tetenbaum, 1975) have suggested that teacher evaluation can be understood as a problem in person perception where the characteristics of the observers (students) can influence their perceptions (evaluations) of the object person (teacher). In fact, Taguiri's (1969) review of the literature suggests that person perception is increasingly viewed as a function of observer, object person, and situational characteristics.

Using person perception theory, Follman (1975) has described two classes of student characteristics: (a) explicit personality factors, and (b) implicit personality factors. Explicit personality factors, such as age, sex, occupation, etc. are generally "more stable, consistent, persistent, and influential than implicit personality characteristics" (Follman, 1975, p. 157). However, Follman suggests that some--if not all--of these explicit factors have generally been unrelated to ratings. Implicit personality factors include "individuals' cognitive style, cognitive controls, field dependence, expectancies, anticipations, and such personality traits as tested in typical personality tests" (Follman, 1975, p. 157). In short, these factors seem highly similar to those of interest in the present investigation. And while the person perception literature less extensively covers implicit personality factors, Follman suggests that many of these factors have been shown to affect person perception.

The present review of the relationship between student character-

istics and ratings suggests that some of these factors--those which are relevant to an educational context--are more likely to relate to ratings than other factors. This has some precedent in person perception theory. For example, Taguiri (1969) has described the selective effect of role relationship on perception:

A person does not see and assess his parents in terms of the same dimensions which he uses for judging his friends; what he looks for in his own children may be different from what he looks for in children in general. Moreover, the inferences about personality drawn from observations differ as a function of the roles of the individuals involved. "Highbrow" behavior in a white man and in a Negro may not lead to the same inference about the person. Bantering behavior in a child leads to one inference when it is directed toward an adult stranger, and another when it is directed toward his own father. (p. 418-419)

This implies that students perceive (evaluate) teachers in their roles as instructors and that the dimensions students use for judging teachers may be different than those used for evaluating other students. If impressions of teachers are formed on the basis of only certain "relevant" information or dimensions, then it may be that student factors more closely associated with these dimensions should exert a greater impact on evaluations than less well associated factors.

The relationship between student characteristics and achievement.

A substantial literature is concerned with instructional strategies which are effective for students with particular characteristics. These studies are primarily concerned with aptitude (or trait)--treatment interactions. For example, researchers may examine the combined effects of a certain student characteristic and particular instructional method on student achievement.

What we know about individual differences leads us to believe that for certain subgroups of individuals, and under certain conditions, reports of overall results misrepresent the true effects. A wide

range of instructional alternatives or treatments needs to be incorporated into research designs in order to determine those treatments that are particularly beneficial for particular subgroups of learners. Such research would be responsive to the question: Given this set of learner characteristics, what is the best way to tailor instruction for this particular type of learner? Research of this kind would compel the educator to consider simultaneously traits, aptitudes, attributes, or characteristics that the learner brings to the situation; the characteristics and dimensions of the instructional agent or medium of instruction; the content to be learned; and the environment in which the instruction takes place. (Berliner & Cahen, 1973, pp. 58-59)

Reviews of this literature are available (Berliner & Cahen, 1973; Cronbach & Snow, 1977). While it is unquestionably true that a variety of instructional strategy by student characteristic interaction effects on achievement have been found, Tobias (1976) notes that the majority of these studies vary instructional rate rather than instructional method (e.g., lecture vs. discussion group vs. tape recording as in Doty, 1967). It might be added that even less attention has been paid to the particular characteristics of the instructor (e.g., style or rapport) which may interact with student characteristics. However, teachers vary not only in their use of different methods of instruction but in their personal style of instruction. Thus, it is of interest to determine the particular characteristics of teachers which are best suited to particular types of students.

Research Questions

TRF researchers may be interested in whether student characteristics explain variability in: (a) student descriptions about their instructor's effectiveness or (b) the relationship between ratings and teacher-

produced achievement. Such evidence may be used to judge TRF validity and generalizability. In addition, the relationship between student characteristics and achievement may be of interest. A review of the literature suggests that student characteristics more readily associated with the educational environment may account for more rating variability than other factors. The theoretical and empirical findings of the person perception literature suggest that student characteristics should relate to ratings, with "relevant" characteristics having a greater relationship to ratings than less "relevant" characteristics.

Several studies have examined the relationship between actual or perceived similarity and ratings. Computed across pooled students, actual similarity scores are difference scores based on student and teacher characteristics. Computed within class, perceived similarity scores are difference scores based on two student characteristics: (a) students' own scores and (b) students' ratings of the instructor's personality. So, rather than examining the amount of rating variability explained by perceived similarity, students' own scores and their ratings of the instructor's personality may be considered separately.

Thus, several research questions have arisen. Generally, explaining variability in: (a) ratings, (b) achievement, and (c) the ratings/achievement relationship is of interest, when each of these criterion measures is considered separately. Students' own OWN characteristics have been mentioned as a source of criterion variability. Of primary interest here is the identification of OWN characteristics which account for criterion variability independent of teacher (TEACHERO effects. Of secondary interest is the identification of the OWN by TEACHER

interactions which account for criterion variability independent of OWN and TEACHER effects. The discussion of perceived similarity led to the introduction of students' perception of instructor (PERCEPT) characteristics as an additional source of criterion variability. Thus, another interest is the identification of PERCEPT characteristics which account for criterion variability independent of OWN characteristics, TEACHER effects, and their interaction. Of final concern are the PERCEPT by TEACHER interactions which account for criterion variability independent of OWN characteristics, TEACHER effects, OWN by TEACHER interactions, and PERCEPT characteristics.

METHOD

Subjects

The subjects were 388 male and female students from the Introductory Psychology course at the University of Manitoba, who volunteered for the experiment without knowing its purpose. These students received credit toward a course requirement for research participation.

Materials

Written materials employed in the study can be found in the Appendix.

Personality Measure

Gough's Adjective Check List (ACL; Gough & Heilbrun, 1967) was used to measure the students' own personality and their perception of the instructor's personality. Scores are available on 24 scales including: number of adjectives checked, unfavorable adjectives checked, defensiveness, favorable adjectives checked, self-confidence, self-control, ability, personal adjustment, achievement, dominance, endurance, order, intraception, nurturance, affiliation, heterosexuality, exhibition, autonomy, aggression, change, succorance, deference, and counseling readiness. Table 2 presents detailed descriptions of the 24 scales. The ACL takes approximately 15-20 minutes to complete. Critiques of the ACL and/or lists of references are available in The Seventh Mental Measurements Yearbook (Buros, 1972), and Personality Tests and Reviews (Buros, 1970). The ACL was chosen for several reasons: (a) it has some established reliability and validity (e.g., Gough & Heilbrun, 1967,

Table 2

Description of the ACL Scales

(Adapted from Gough & Heilbrun, 1967 pp. 7-11)

Total Checked

The individual high on this variable tends to be described as emotional, adventurous, wholesome, conservative, enthusiastic, unintelligent, frank, and helpful. He is active, apparently means well, but tends to blunder. The man with low scores tends more often to be quiet and reserved, more tentative and cautious in his approach to problems, and perhaps at times unduly tactiturn and aloof. He is more apt to think originally and inventively, but is perhaps less effective in getting things done.

Defensiveness

The higher-scoring person is apt to be self-controlled and resolute in both attitude and behavior, and insistent and even stubborn in seeking his objectives. His persistence is more admirable than attractive. The lower-scoring subject tends to be anxious and apprehensive, critical of himself and others, and given to complaints about his circumstances. He not only has more problems than his peers, but tends to dwell on them and put them at the center of his attention

Favorable

The high-scoring individual appears to be motivated by a strong desire to do well and to impress others, but always by virtue of hard work and conventional endeavour. The reaction of others is to see him as dependable, steady, conscientious, mannerly, and serious; there is also the suspicion that he may be too concerned about others, and lacking in verve and quickness of mind. The low-scoring subject is much more of an individualist--more often seen as clever, sharp-witted, headstrong, pleasure-seeking, and original in thought and behavior. His emotions being more accessible, he also more often experiences anxiety, self-doubts, and perplexities.

Unfavorable

The high-scoring subject strikes others as rebellious, arrogant, careless, conceited, and cynical. He tends to be a disbeliever, a skeptic, and a threat to the complacent beliefs and attitudes of his fellows. The low-scorer is more placid, more obliging, more mannerly, more tactful, and probably less intelligent.

(Table 2 - continued)

Self-Confidence

The high-scorer is assertive, affiliative, outgoing, persistent, an actionist. He wants to get things done, and is impatient with people or things standing in his way. He is concerned about creating a good impression, and is not above cutting a few corners to achieve this objective. He makes a distinct impression on others, who see him as forceful, self-confident, determined, ambitious, and opportunistic. The low-scoring person is a much less effective person in the everyday sense of the word--he has difficulty in mobilizing himself and taking action, preferring inaction and contemplation. Others see him as unassuming, forgetful, mild, preoccupied, reserved, and retiring.

Self-Control

High scorers tend to be serious, sober individuals, interested in and responsive to their obligations. They are seen as diligent, practical, and loyal workers. At the same time there may be an element of over-control, too much emphasis on the proper means for attaining the ends of social living. At the other end of the scale one seems to find the inadequately socialized person, headstrong, irresponsible, complaining, disorderly, narcissistic, and impulsive. Needless to say, the low-scoring subject tends to be described in unflattering terms, even including such words as obnoxious, autocratic, and thankless.

Liability

The high-scoring subject is seen favorably as spontaneous, but unfavorably as excitable, temperamental, restless, nervous, and high-strung. The psychological equilibrium, the balance of forces, is an uneasy one in this person and he seems impelled toward change and new experience in an endless flight from his perplexities. The low-scorer is more phlegmatic, routinized, planful, and conventional. He reports stricter opinions on right and wrong practices, and a greater need for order and regularity. He is described by observers as thorough, organized, steady, and unemotional.

Personal Adjustment

The high-scoring subject is seen as dependable, peaceable, trusting, friendly, practical, loyal, and wholesome. He fits in well, asks for little, treats others with courtesy, and works enterprisingly toward his own goals. He may or may not understand himself psychodynamically, but he nonetheless seems to possess the capacity to "love and work." The subject low on the

(Table 2 - continued)

adjustment scale sees himself as at odds with other people and as moody and dissatisfied. This view is reciprocated by observers, who describe the low scorer as aloof, defensive, anxious, inhibited, worrying, withdrawn, and unfriendly. What appears to begin as a problem in self-definition eventuates as a problem in inter-personal living.

Achievement

The high-scoring subject on "achievement" is usually seen as intelligent and hard-working, but also as involved in his intellectual and other endeavours. He is determined to do well and usually succeeds. His motives are internal and goal-centered rather than competitive, and in his dealings with others he may actually be unduly trusting and optimistic. The low-scoring subject on "achievement" is more skeptical, more dubious about the rewards which might come from effort and involvement, and uncertain about risking his labors. He tends to be somewhat withdrawn and dissatisfied with his current status.

Dominance

The high-scorer on this scale is a forceful, strong-willed and persevering individual. He is confident of his ability to do what he wishes and is direct and forthright in his behavior. The low scorer on "dominance" is unsure of himself, and indifferent to both the demands and the challenges of interpersonal life. He stays out of the limelight, and avoids situations calling for choice and decision-making.

Endurance

The subject on "endurance" is typically self-controlled and responsible, but also idealistic and concerned about truth and justice. By nature conventional, he may nonetheless (because of his sense of rectitude) find himself championing unconventional ideas and unpopular causes. The low-scorer on "endurance" on the other hand, is erratic and impatient, intolerant of prolonged effort or attention, and apt to change in an abrupt and quixotic manner.

Order

High-scorers on "order" are usually sincere and dependable but at the cost of individuality and spontaneity. These self-denying and inhibitory trends may actually interfere with the attainment of the harmony and psychic order which they seek. Low-scorers are quicker in temperament and reaction, and might often be called impulsive. They prefer complexity and variety, and dislike delay, caution, and deliberation.

(Table 2 - continued)

Intracception

The high-scorer on "intracception" is reflective and serious, as would be expected; he is also capable, conscientious, and knowledgeable. His intellectual talents are excellent and he derives pleasure from their exercise. The low-scorer may also have talent, but he tends toward profligacy and intemperateness in its use. He is aggressive in manner, and quickly becomes bored or impatient with any situation where direct action is not possible. He is a doer, not a thinker.

Nurturance

The subject high on this scale is of a helpful, nurturant disposition, but sometimes too bland and self-disciplined. His dependability and benevolence are worthy qualities, but he may nonetheless be too conventional and solicitous of the other person. The subject scoring low on "nurturance" is the opposite: skeptical, clever, and acute, but too self-centered and too little attentive to the feelings and wishes of others.

Affiliation

The high-scorer on "affiliation" is adaptable and anxious to please, but not necessarily because of altruistic motives; i.e., he is ambitious and concerned with position, and may tend to exploit others and his relationships with them in order to gain his ends. The low-scorer is more individualistic and strong-willed, though perhaps not out of inner resourcefulness and independence. He tends to be less trusting, more pessimistic about life, and restless in any situation which intensifies or prolongs his contacts with others.

Heterosexuality

The high-scorer on "heterosexuality" is interested in the opposite sex as he is interested in life, experience, and most things around him in a healthy, direct, and outgoing manner. He may even be a bit naive in the friendly ingenuousness in which he approaches others. The low-scorer thinks too much, as it were, and dampens his vitality; he tends to be dispirited, inhibited, shrewd, and calculating in his interpersonal relationships.

Exhibition

Persons who are high on this scale tend to be self-centered and even narcissistic. They are poised, self-assured, and able to meet situations with aplomb, but at the same time they are quick tempered and irritable. In their dealings with others they

(Table 2 - continued)

are apt to be opportunistic and manipulative. Persons who score low tend toward apathy, self-doubt, and undue inhibition of impulse. They lack confidence in themselves and shrink from any encounter in which they will be visible or "on stage."

Autonomy

The high-scorer on "autonomy" is independent and autonomous, but also assertive and self-willed. He tends to be indifferent to the feelings of others and heedless of their preferences when he himself wishes to act. The low-scorer is of a moderate and even subdued disposition. He hesitates to take the initiative, preferring to wait and follow the dictates of others.

Aggression

The individual high on this scale is both competitive and aggressive. He seeks to win, to vanquish, and views others as rivals. His impulses are strong, and often under-controlled. In an appropriate situation he may drive on to worthy attainment, but often his behaviors will be self-aggrandizing and disruptive. The individual who is low on "aggression" is much more of a conformist, but not necessarily lacking in courage or tenacity. He tends to be patiently diligent, and sincere in his relationships with others.

Change

Persons high on "change" are typically perceptive, alert, and spontaneous individuals who comprehend problems and situations rapidly and incisively and who take pleasure in change and variety. They have confidence in themselves and welcome the challenges to be found in disorder and complexity. The low-scorer seeks stability and continuity in his environment, and is apprehensive of ill-defined and risk-involving situations. In temperament he is patient and obliging, concerned about others, but lacking in verve and energy.

Succorance

"Succorance" appears to depict, at its high end, a personality which is trusting, guileless, and even naive in its faith in the integrity and benevolence of others. The high-scorer is dependent on others, seeks support, and expect to find it. The low-scorer, on the contrary, is independent, resourceful, and self-sufficient, but at the same time prudent and circumspect. He has a sort of quiet confidence in his own worth and capability.

(Table 2 - continued)

Abasement

High-scorers on "abasement" are not only submissive and self-effacing, but also appear to have problems on self-acceptance. They see themselves as weak and undeserving, and face the world with anxiety and foreboding. Their behavior is often self-punishing, perhaps in the hope of forestalling criticism and rejection from without. The low-scorer is optimistic, poised, productive, and decisive. Not fearing others, he is alert and responsive to them. His tempo is brisk, his manner confident, and his behavior effective.

Deference

The individual scoring high on "deference" is typically conscientious, dependable, and persevering. He is self-denying not so much out of any fear of others or inferiority to them as out of a preference for anonymity and freedom from stress and external demands. He attends modestly to his affairs, seeking little, and yielding always to any reasonable claim by another. The individual with a low score on "deference" is more energetic, spontaneous, and independent; he likes attention, like to supervise and direct others, and to express his will. He is also ambitious, and is not above taking advantage of others and coercing them if he can attain a goal in so doing.

Counseling Readiness

The high-scorer on "counseling readiness" is predominantly worried about himself and ambivalent about his status. He feels left out of things, unable to enjoy life to the full, and unduly anxious. He tends to be preoccupied with his problems and pessimistic about his ability to resolve them constructively. The low-scorer is more or less free of these concerns. He is self-confident, poised, sure of himself and outgoing. He seeks the company of others, likes activity, and enjoys life in an uncomplicated way.

report an average test-retest reliability over 10 weeks for the scales of .72 and significant relationships of the ACL with behavioural measures and other personality tests); (b) it contains some scales which appear relevant to an educational environment; (c) it can be used to measure both the student's own characteristics as well as the student's perception of the instructor's qualities; and (d) It can be administered to a group of students and is not overly time-consuming to complete.

Videotaped Lectures

Four colour, videotaped lectures were constructed which differ systematically with respect to teaching expressiveness (low, high) and lecture content (low, high). The topic of all lectures was sex roles. The high content lecture was prepared from actual lecture notes which contained all the information that was to be presented. The low content lecture was prepared by eliminating a number of teaching points contained in the high content lecture, using a procedure designed to: (a) maintain script logic and coherence; (b) distribute the remaining teaching points evenly throughout the script, and (c) ensure that the importance of the remaining and eliminated teaching points were approximately equal. The filler information that was used in place of the omitted teaching points included: (a) discussion of unrelated examples, events, etc., (b) repeated emphasis on what information was going to be covered without such coverage, and (c) circular discussion of unrelated concepts and ideas. This filler ensured that all videotapes were approximately 25 minutes long.

All lectures were delivered by a tenured professor in the Psychology

Department, not teaching Introductory Psychology, who had a reputation among his colleagues and throughout the university community as an outstanding instructor. For the high expressiveness lectures, the professor was asked to role play the following characteristics: enthusiasm, humor, friendliness, physical movement, vocal inflection, and charisma. For the low expressiveness lectures, the professor was asked to minimize his use of these characteristics. The making of these tapes was guided by previous experience with the black and white tapes constructed by Ware and Williams (1975), and Perry, Abrami, and Leventhal (1978), who also varied instructor expressiveness and lecture content. Perry, Leventhal, and Abrami (Note 5) found that these colour tapes produced similar ratings effects when compared to the Ware and Williams (1975) tapes.

Dependent Measures

A 30-item multiple choice achievement test (ACH) was used to assess how much students learned from the lecture. This quiz format was selected because of its frequent use as a criterion measure in the large undergraduate survey courses which are generally studied in TRF research. Perry, Leventhal, and Abrami (Note 5) found a split-half reliability for the quiz of .82 using Cronbach's alpha. A 44-item TRF assessed student evaluations of the instructor. This TRF consists of (a) an overall teaching ability item similar to that employed by Sullivan and Skanes (1974), (b) 25-items previously employed by Leventhal, Perry, and Abrami (1977) which were adapted for use with videotaped lectures from a TRF developed by Hildebrand and Wilson (1970), and (c) 18-items previously employed by Ware and Williams (1975) from a TRF developed

by Pohlmann (1975).

Procedure

All students were instructed that they were participating in research on teacher and student factors which affect student satisfaction with, and performance in, a college level course. Immediately prior to viewing a videotaped lecture, all students completed the ACL for themselves. Students then: (a) viewed one of four videotaped lectures, (b) completed the TRF and ACH, and (c) completed the ACL for the instructor. At the end of the experiment interested students were debriefed as to the true nature of the experiment, during which time the experimental design and hypotheses were outlined.

Data Analysis

For each criterion measure, the data were analyzed via stepwise multiple linear regression analysis (MRA) where the categorical teacher variables (style and content) and their interaction were dummy variable coded (see Kerlinger & Pedhazur, 1973). The MRA was run using the regression subprogram from SPSS (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975). In MRA, the order in which the predictor variables are introduced into the equation is critical when intercorrelations exist. This order determines how the variance due to regression will be divided among the predictors. In the present investigation, the order of entry of sets of predictors was specified beforehand. The set of TEACHER characteristics (style, content, and style by content) was entered first, the set of OWN characteristics second, the OWN by TEACHER interactions third, the

PERCEPT characteristics fourth, and the PERCEPT by TEACHER interactions last. To estimate the variance accounted for by the set of OWN characteristics independent of TEACHER characteristics, the increment in variance accounted for from step one to step two was noted. Likewise, to estimate the variance accounted for by the set of OWN by TEACHER interactions, the increment in variance accounted for from step two to step three was noted. This a priori ordering approach partitions all of the explainable rating variance. It does not allow one to determine the unique contribution of each set of predictors.

If a set of characteristics, other than the set of TEACHER characteristics, resulted in a significant increment in rating variance accounted for, the minimum number of characteristics necessary to account for much of the variance explained by the total set was determined. This was also accomplished via MRA. For example, imagine that the set of OWN characteristics resulted in a significant increment in rating variance explained. After entering the TEACHER characteristics into the MRA, each OWN characteristic may be included in a forward (stepwise) manner (SPSS, 1975). The order of inclusion of each OWN characteristic is determined by the respective contribution of each characteristic to explained variance. The best subset of characteristics is found when the inclusion of another variable in the equation fails to account for a meaningful increment in rating variance.

The MRA approach can be used in the situation where there are 24 continuous student variables (i.e., the scales of the ACL) and two teacher variables. Excluding all the interactions among student characteristics, there would still remain 26 main effects, 49 two-way interactions, and

24 three-way interactions! But as Kerlinger and Pedhazur (1973, pp. 441-444) suggest, the use of a large number of predictor variables, especially when the variables are intercorrelated, can yield regression weights which are highly unstable. Although the stability of the regression weights can be improved by increasing sample size, the number of students necessary to substantially improve reliability for 99 predictors is in excess of the total population of the University of Manitoba subject pool!

An alternative approach is to reduce the number of predictors through theory and factor analysis. Accordingly, the ACL scores for all students were factor analyzed by orthogonal varimax rotation using SPSS (1975) to find a subset of scales where several criteria were considered: (a) the scale loaded highly on one factor only; (b) the scale was a fairly reliable measure; and (c) the scale appeared relevant to an educational context. These few scales, rather than the 24 scales, were used in the MRA.

The 44-item TRF employed in this investigation can provide at least three student rating scores. Following Leventhal et al. (1977), and Ware and Williams (1975), these scores were: (a) the response to the overall ability item, (b) the mean of the Hildebrand and Wilson (1970) items, and (c) the mean of the Pohlmann (1975) items. These scores may be related, both conceptually and statistically. Thus, it might seem appropriate to employ a data analytic technique which permits analysis of the set of dependent measures. Multivariate multiple regression, or canonical correlation analysis (CCA) is such a

technique.⁴

Using CCA suggests that the researcher is especially interested in the criterion variables as a set. While the effects of teacher and student variables on the set of student rating scores is of some interest, of greater interest is the relationship of the predictor variables to each student rating score since using more than one TRF is rare in applied settings. Thus, the CCA was not run.

Investigating the ratings/achievement relationship posed two special problems. The first problem was how to probe the ratings/achievement relationship for only four teacher conditions using MRA. In a prior study, Leventhal et al. (1977) investigated whether teacher effects on ratings and achievement vary as a function of student characteristics by using a repeated measures analysis of variance (ANOVA). Each rating and achievement raw score was transformed into a z-score based on the sample mean and variance of the measure. The variance for a measure was computed for each treatment group separately and pooled to exclude variation from group treatment effects. The logic of this ANOVA procedure was extended to the present situation where there are continuous predictor variables in the design by using MRA to analyze

⁴Discussions of CCA can be found in Cooley and Lohnes (1971), Harris (1975), and Kerlinger and Pedhazur (1973). In CCA, the canonical correlations (there can be more than one) represent the maximum correlations possible between the set of predictor variables and the set of criterion variables. In order to employ a stepwise procedure, however, it is necessary to find a composite score based on the canonical coefficients for the criterion variables when all the predictor variables are in the equation. Then a stepwise MRA can be run using this composite score as the criterion. Stepwise CCA is generally not recommended because the weights assigned to predictor and criterion variables might change from step-to-step. Thus, changes in the results from step-to-step might be difficult to interpret.

repeated measures data (see Kerlinger & Pedhazur, 1973, pp. 218-221). For the tests of within subjects effects, this normally requires the generation of $N - 1$ vectors that identify the subjects. However, Pedhazur (1977) has recently suggested a method of exploring the within subjects effects which require only a single vector to code subjects.

A second problem was how to transform the raw scores to standard scores when some predictor variables are continuous. When the predictor variables are only categorical, Gabriel and Hopkins (1974) suggest that the transformation should be based on the sample mean and variance pooled across groups. The variance for each group is calculated separately and then pooled to avoid confounding with group differences in means. When the predictor variables are also continuous, however, it is not possible to divide subjects into a few groups and calculate separate within group variances. However, it is possible to compute the total within group variance, or the amount of unexplained variance, in the criterion measure. The error term used in an F-test when all the predictors have been entered into an MRA, or the variance of estimate (Kerlinger & Pedhazur, 1973, p. 66), is an index of the variance in the criterion which is not "confounded" by the variance explained by the predictors. The variance of estimate can thus be used to transform raw scores to standard scores.⁵

⁵In developing this approach to the transformation of raw scores, the author benefited from discussions with R. Gabriel, October, 1977.

RESULTS

Factor Analysis

The factor analysis employing orthogonal varimax rotation was run using subprogram FACTOR from SPSS (1975). The initial factor matrix solution based on the principal components method (i.e., unities in the main diagonal of the correlation matrix) revealed that four factors could jointly account for 84.8% (38.2%, 26.6%, 12.2%, and 7.8%, respectively) of the variance in the ACL. The orthogonal varimax rotated four-factor matrix (i.e., communality estimates in the main diagonal of the correlation matrix) is presented in Table 3. The large number of ACL scales that loaded highly on Factor 1 suggest that it is a fairly general factor. The Deference (DEF) scale was chosen to represent Factor 1 because it loaded high on that factor only and is reasonably reliable (.80 test-retest; see Gough & Heilbrun, 1965). Factor 2 also had several scales with high loadings. None of the scales with high loadings on Factor 2 were substantially independent of all the remaining factors. The Affiliation (AFF) scale was chosen to represent Factor 2 because it loaded highest on this factor and is fairly reliable (.83). The Achievement (ACH) scale was chosen to represent Factor 3 because of its high loading on this factor, reasonable reliability (.78), and its apparent relationship to an educational context. Both the Order and Intraception scales high higher loadings on Factor 3 than ACH and were practically independent of the other factors. However, these scales were less reliable than ACH (.60 for Order and .59 for Interception) and most importantly did not seem as related to an educational context. Of the two

TABLE 3

Orthogonal Varimax Rotated Factor Matrix

<u>ACL Scale</u>	<u>Factor 1</u>	<u>Factor 2</u>	<u>Factor 3</u>	<u>Factor 4</u>
Total checked	0.048	0.603	0.374	0.689
Defensiveness	0.162	0.656	0.616	0.057
Favorable	0.089	0.786	0.585	0.050
Unfavorable	0.078	-0.011	-0.142	<u>0.950</u>
Self-Confidence	0.762	0.341	0.403	0.015
Self-Control	-0.572	0.105	0.606	-0.265
Liability	0.259	0.700	-0.076	0.214
Personal Adjustment	-0.122	0.552	0.579	-0.290
Achievement	0.462	0.310	<u>0.737</u>	-0.087
Dominance	0.762	0.258	0.485	-0.216
Endurance	0.058	0.070	0.916	-0.151
Order	-0.067	0.099	0.919	0.049
Intracception	-0.184	0.613	0.583	-0.069
Nurturance	-0.356	0.716	0.237	-0.401
Affiliation	0.077	<u>0.844</u>	0.409	0.090
Heterosexuality	0.231	0.791	0.015	-0.026
Exhibition	0.822	0.276	-0.115	0.016
Autonomy	0.847	-0.014	-0.056	0.272
Aggression	0.743	-0.390	-0.313	0.268
Change	0.579	0.463	-0.209	0.022
Succorance	-0.496	-0.013	-0.343	0.563
Abasement	-0.869	0.014	-0.203	0.233
Deference	<u>-0.925</u>	0.085	0.075	-0.055
Counseling Readiness	-0.716	-0.499	-0.153	0.233

scales with high loadings on Factor 4, Unfavorable (UNFAV) was selected because of its independence of the remaining factors and good reliability (.81).

Ratings

Means (\bar{X} 's) and standard deviations (SD's) of the predictors and ratings criteria employed in the stepwise MRAs are presented in Table 4. For all three rating criteria, the sets of predictors were entered into the MRAs in the following order: TEACHER characteristics (T), OWN characteristics (O), T x O, PERCEPT characteristics (P), and T x P. Thus, for any criterion, the variance explained by the set of TEACHER characteristics includes the unique variance due to T and the variance in the criterion which T shares with all the other sets of predictors. The variance explained by O is independent of T but not any of the other sets. Only the variance explained by the set entered last (T x P) is orthogonal to the other sets. Kerlinger and Pedhazur (1973, p. 286) argue against judging the meaningfulness of a predictor/criterion relationship only in terms of a level of statistical significance because a predictor may account for a significant--yet small--proportion of criterion variance. In the present study, a set of characteristics will be considered meaningfully related to the criterion if that relationship is: (a) significant at the .01 level and (b) accounts for at least 5% of the variance.

A summary of the regression of scores on the Sullivan and Skanes item is presented in Table 5. As expected, the set of TEACHER characteristics accounted for a significant amount of rating variance, $F(3, 352)$

TABLE 4

Means and Standard Deviations of Predictor and Criterion Measures Used in the MRAs

Measure	\bar{X}	SD	Measure	\bar{X}	SD
<u>TEACHER (T)</u>			<u>PERCEPT (P)</u>		
Style (S)	0.021	1.000	Percept Unfavorable (PU)	8.077	7.816
Content (C)	-0.026	1.000	Percept Achievement (PACH)	4.887	7.573
S x C	-0.036	1.000	Percept Affiliation (PAFF)	8.144	8.143
<u>OWN (O)</u>			Percept Deference (PD)	1.539	4.609
Own Unfavorable (OU)	11.296	8.711	<u>T x P</u>		
Own Achievement (OACH)	10.059	5.929	S x PU	2.603	10.937
Own Affiliation (OAFF)	20.644	6.478	S x PACH	-5.469	7.163
Own Deference (OD)	1.933	5.880	S x PAFF	-5.273	10.241
<u>T x O</u>			S x PD	2.276	4.293
S x OU	0.575	14.259	C x PU	0.052	11.243
S x OACH	0.570	11.668	C x PACH	-0.562	8.997
S x OAFF	0.180	21.649	C x PAFF	-1.052	11.472
S x OD	-0.866	6.129	C x PD	-0.482	4.836
C x OU	-0.936	14.240	S x C x PU	-1.217	11.177
C x OACH	-0.740	11.658	S x C x PACH	0.299	9.009
C x OAFF	-0.644	21.640	S x C x PAFF	0.469	11.511
C x OD	0.717	6.148	S x C x PD	0.244	4.854
S x C x OU	-1.451	14.197	<u>MEASURES (M)</u>		
S x C x OACH	0.059	11.682		0.000	1.001
S x C x OAFF	-0.613	21.641	<u>M x T</u>		
S x C x OD	-0.134	6.188	M x S	0.000	1.001
			M x C	0.000	1.001
			M x S x C	0.000	1.001

Table 4 - Continued

Measure	\bar{X}	SD	Measure	\bar{X}	SD
<u>M x O</u>			<u>M x T x P</u>		
M x OU	0.000	14.271	M x S x PU	0.000	11.243
M x OACH	0.000	11.682	M x S x PACH	0.000	9.014
M x OAFF	0.000	21.650	M x S x PAFF	0.000	11.521
M x OD	0.000	6.190	M x S x PD	0.000	4.860
<u>M x T x O</u>			M x C x PU	0.000	11.243
M x S x OU	0.000	14.271	M x C x PACH	0.000	9.014
M x S x OACH	0.000	11.682	M x C x PAFF	0.000	11.521
M x S x OAFF	0.000	21.650	M x C x PD	0.000	4.860
M x S x OD	0.000	6.190	M x S x C x PU	0.000	11.243
M x C x OU	0.000	14.271	M x S x C x PACH	0.000	9.014
M x C x OACH	0.000	11.682	M x S x C x PAFF	0.000	11.521
M x C x OAFF	0.000	21.650	M x S x C x PD	0.000	4.860
M x C x OD	0.000	6.190	<u>CRITERIA</u>		
M x S x C x OU	0.000	14.271	Sullivan & Skanes (SS)	2.273	1.163
M x S x C x OACH	0.000	11.682	Hildebrand & Wilson (HW)	2.516	0.773
M x S x C x OAFF	0.000	21.650	Pohlmann (PN)	2.601	0.980
M x S x C x OD	0.000	6.190	Achievement (ACH)	16.972	5.583
<u>M x P</u>			SS/ACH	0.000	1.617
M x PU	0.000	11.243	HW/ACH	0.000	3.254
M x PACH	0.000	9.014	PN/ACH	0.000	3.163
M x PAFF	0.000	11.521			
M x PD	0.000	4.860			

TABLE 5
Summary of Regression of TRFs and Achievement

	Source	SS	df	MS	F	p	Change in R ²
SULLIVAN & SKANES	TEACHER (T)	259.488	3	86.496	168.934	.001	.496
	OWN (O)	0.983	4	0.246	0.480	.750	.002
	T x O	4.026	12	0.336	0.656	.793	.008
	PERCEPT (P)	70.746	4	17.687	34.545	.001	.135
	T x P	7.477	12	0.623	1.217	.269	.014
	Error	180.321	352	0.512			
HILDEBRAND & WILSON	TEACHER (T)	136.929	3	45.643	271.685	.001	.592
	OWN (O)	1.737	4	0.434	2.583	.037	.008
	T x O	3.089	12	0.257	1.530	.111	.013
	PERCEPT (P)	27.737	4	6.934	41.274	.001	.120
	T x P	2.739	12	0.228	1.357	.185	.012
	Error	59.241	352	0.168			
POHLMANN	TEACHER (T)	236.473	3	78.824	359.927	.001	.636
	OWN (O)	0.993	4	0.248	1.132	.341	.003
	T x O	3.890	12	0.324	1.479	.130	.010
	PERCEPT (P)	46.274	4	11.569	52.826	.001	.124
	T x P	7.001	12	0.583	2.662	.002	.019
	Error	77.248	352	0.219			
Achievement	TEACHER (T)	3709.685	3	1236.562	64.374	.001	.308
	OWN (O)	200.685	4	50.171	2.612	.035	.017
	T x O	272.609	12	22.718	1.183	.294	.023
	PERCEPT (P)	335.168	4	83.792	4.362	.002	.028
	T x P	781.112	12	65.093	3.389	.001	.065
	Error	6761.430	352	19.209			

= 168.934, $p = .001$. The only other set of characteristics to account for a significant amount of rating variance was the PERCEPT characteristics, $F(4, 352) = 34.545$, $p = .001$. PERCEPT characteristics accounted for 13.5% of the rating variance independent of T, 0, and T x 0.

The results for the mean of the Hildebrand and Wilson items and the Pohlmann items were similar to those for the Sullivan and Skanes item (see Table 5). The set of TEACHER characteristics was significantly related to the Hildebrand and Wilson TRF, $F(3, 352) = 271.685$, $p = .001$. Otherwise, only the set of PERCEPT characteristics was significantly related to the Hildebrand and Wilson criterion, $F(4, 352) = 41.274$, $p = .001$, and accounted for more than 5% of the variance (12%). The set of TEACHER characteristics was significantly related to the Pohlmann TRF, $F(3, 352) = 359.927$, $p = .001$. In addition, the set of T x P characteristics was significantly related to the Pohlmann criterion, $F(12, 352) = 2.662$, $p = .002$, but accounted for under 2% of the variance. The set of PERCEPT characteristics was significantly related to the criterion, $F(4, 352) = 52.826$, $p = .001$, and accounted for 12.4% of the Pohlmann variance.

Thus, for all three rating criteria, the only set of characteristics, other than TEACHER characteristics, to account for a meaningful amount of rating variance was the PERCEPT characteristics. In each case, the set accounted for more than 10% of the variance. In addition, the variance explained by the set of PERCEPT characteristics was independent of the influence of all other sets except the set of T x P interactions.

Subsequent analyses found the minimum number of PERCEPT character-

istics necessary to account for a large percentage of the rating variance explained by the total set. After entering the T, 0, and T x 0 sets of characteristics, each PERCEPT characteristic was entered in the analyses in a forward (stepwise) manner. For each rating criterion, the results were highly consistent. As shown in Table 6, students' perceptions of instructors' need for achievement (PACH) was always the first characteristic to enter the equations, was significantly related to each criterion, and generally accounted for 80% of the rating variance explained by the entire set.

Achievement

The computation of the regression of the achievement scores was similar to ratings. Four sets of predictors (T, 0, T x 0, P, T x P) were entered in a priori order. The summary of this regression is shown at the bottom of Table 5. Aside from the set of TEACHER characteristics, $F(3, 352) = 64.374$, $p = .001$, only the set of TEACHER x PERCEPT characteristics was significantly related to achievement, $F(12, 352) = 3.389$, $p = .001$, and accounted for more than 5% of the achievement variance (6.5%). Since this set entered last in the regression equation, it accounted for achievement variance independent of all other sets.

After entering all other sets, individual characteristics from the T x P set were entered in an MRA in a forward (stepwise) manner. As shown in the bottom of Table 6, the first variable, teacher style by students' perceptions of instructors' deference (S x PD), was significantly related to achievement, $F(1, 352) = 32.441$, $p = .001$. S x PD also accounted for 80% of the achievement variance explained by the T x P set.

TABLE 6

Forward (Stepwise) MRA to Find Best Subset of Predictors

	Source	SS	df	MS	F	p	Change in R ²	Proportion of Set
SULLIVAN & SKANES	PERCEPT (P)	70.746	4	17.687	34.545	.001	.135	1.000
	PACH	57.734	1	57.734	112.762	.001	.110	0.815
	PU	6.613	1	6.613	12.916	.001	.013	0.096
	PAFF	6.371	1	6.371	12.443	.001	.012	0.089
	PD	0.028	1	0.028	0.055	.815	.000	0.000
	Error	180.321	352	0.512				
HILDEBRAND & WILSON	PERCEPT (P)	27.737	4	6.934	41.274	.001	.120	1.000
	PACH	23.569	1	23.569	140.292	.001	.102	0.850
	PAFF	2.216	1	2.216	13.190	.001	.010	0.083
	PU	1.604	1	1.604	9.548	.002	.007	0.058
	PD	0.348	1	0.348	2.071	.151	.002	0.017
	Error	59.241	352	0.168				
POHLMANN	PERCEPT (P)	46.274	4	11.569	52.826	.001	.124	1.000
	PACH	36.817	1	36.817	168.114	.001	.099	0.798
	PU	7.134	1	7.134	32.575	.001	.019	0.153
	PAFF	2.169	1	2.169	9.904	.002	.006	0.048
	PD	0.154	1	0.154	0.703	.402	.000	0.000
	Error	77.248	352	0.219				
Achievement	T x P	781.112	12	65.093	3.389	.001	.065	1.000
	S x PD	623.160	1	623.160	32.441	.001	.052	0.800
	S x PAFF	39.972	1	39.972	2.081	.150	.003	0.046
	S x C x PACH	28.819	1	28.819	1.500	.221	.002	0.031
	S x C x PD	35.282	1	35.282	1.837	.176	.003	0.046
	C x PD	16.112	1	16.112	0.839	.360	.001	0.015
	C x PACH	9.153	1	9.153	0.476	.491	.001	0.015
	C x PAFF	21.761	1	21.761	1.133	.288	.002	0.031
	S x PU	6.012	1	6.012	0.313	.576	.001	0.015
	S x C x PU	0.675	1	0.675	0.035	.852	.000	0.000
	S x PACH	0.149	1	0.149	0.008	.929	.000	0.000
	S x C x PAFF	0.015	1	0.015	0.001	.975	.000	0.000
Error	6761.430	352	19.209					

The Ratings/Achievement Relationship

Means and standard deviations of the predictors (including the within subject factors) and ratings/achievement criteria employed in the repeated measures MRAs are presented in Table 4. For the three MRAs, the criterion measure consisted of rating scores (i.e., Sullivan & Skanes, Hildebrand & Wilson, or Pohlmann) and achievement scores converted to z-scores as previously described. For each MRA, the sets of predictors were entered in the following order: TEACHER characteristics (T), OWN characteristics (O), T x O, PERCEPT characteristics (P), T x P, MEASURES (M), M x T, M x O, M x T x O, M x P, and M x T x P. While tests of the between subject factors (i.e., T, O, T x O, P, and T x P) were not independent of each other, tests of the within subject factors (i.e., M through M x T x P) were orthogonal to each other and the between subject factors.

A summary of the repeated measures regression of the transformed Sullivan and Skanes/achievement scores is presented in Table 7. Other than the set of TEACHER characteristics [$F(3, 352) = 189.072, p = .001$] and the MEASURES by TEACHER interactions [$F(3, 352) = 142.322, p = .001$], only two sets were significant and accounted for at least 5% of the variance. The set of PERCEPT characteristics accounted for a significant amount of ratings/achievement variance, $F(4, 352) = 35.282, p = .001$, which was 7.1% of criterion variance. The set of MEASURES by PERCEPT interactions was significant, $F(4, 352) = 32.146, p = .001$, accounting for 6.2% of the variance. Taken together, these findings suggest that PERCEPT characteristics explain some of the covariation between ratings and achievement and also explain some of the variation in ratings which

TABLE 7
Summary of Repeated Measures MRAs

	Source	SS	df	MS	F	p	Change in R ²
SULLIVAN & SKANES	TEACHERS (T)	579.128	3	193.043	189.072	.001	.286
	OWN (O)	2.199	4	0.550	0.539	.707	.001
	T x O	7.693	12	0.641	0.628	.819	.004
	PERCEPT (P)	144.089	4	36.023	35.282	.001	.071
	T x P	18.837	12	1.570	1.538	.109	.009
	Error	359.302	352	1.021			
	MEASURES (M)	0.0	1	0.0	-	-	
	M x T	419.710	3	139.903	142.322	.001	.207
	M x O	2.094	4	0.524	0.533	.712	.001
	M x T x O	10.478	12	0.873	0.888	.559	.004
	M x P	126.400	4	31.600	32.146	.001	.062
	M x T x P	11.771	12	0.981	0.998	.450	.006
	Error	346.143	352	0.983			
HILDEBRAND & WILSON	TEACHERS (T)	2572.574	3	857.525	280.604	.001	.314
	OWN (O)	27.415	4	6.854	2.243	.064	.003
	T x O	52.797	12	4.400	1.440	.145	.006
	PERCEPT (P)	508.749	4	127.187	41.619	.001	.062
	T x P	56.031	12	4.669	1.528	.112	.007
	Error	1075.708	352	3.056			
	MEASURES (M)	0.0	1	0.0	-	-	
	M x T	2271.715	3	757.238	257.739	.001	.277
	M x O	34.446	4	8.612	2.931	.021	.004
	M x T x O	56.984	12	4.749	1.616	.085	.007
	M x P	528.417	4	132.104	44.964	.001	.057
	M x T x P	42.764	12	3.564	1.213	.272	.005
	Error	1034.100	352	2.938			
POHLMANN	TEACHERS (T)	2637.006	3	879.002	371.043	.001	.340
	OWN (O)	10.383	4	2.596	1.096	.358	.001
	T x O	40.176	12	3.348	1.413	.157	.005
	PERCEPT (P)	500.301	4	125.075	52.797	.001	.065
	T x P	82.268	12	6.856	2.894	.001	.029
	Error	833.989	352	2.369			
	MEASURES (M)	0.0	1	0.0	-	-	
	M x T	2283.370	3	761.123	339.787	.001	.294
	M x O	10.783	4	2.696	1.204	.309	.001
	M x T x O	41.349	12	3.446	1.538	.108	.005
	M x P	461.462	4	115.366	51.503	.001	.060
	M x T x P	65.232	12	5.436	2.427	.005	.008
	Error	788.374	352	2.240			

is independent of achievement (and vice-versa).

A summary of the repeated measures regression of the remaining transformed ratings (i.e., Hildebrand & Wilson, Pohlmann)/achievement scores are also presented in Table 7. These results were similar to the prior repeated measures regression. The set of PERCEPT characteristics accounted for a significant amount of Hildebrand and Wilson/achievement variance, $F(4, 352) = 41.619$, $p = .001$, and Pohlmann/achievement variance, $F(4, 352) = 52.797$, which was 6.2% and 6.5%, respectively, of the criterion variance. The set of MEASURES by PERCEPT interactions was significant, $F(4, 352) = 44.964$, $p = .001$, accounting for 5.7% of the Hildebrand and Wilson/achievement variance, while accounting for 6.0% of the Pohlmann/achievement variance significantly, $F(4, 352) = 51.503$, $p = .001$.

Thus, for all three rating/achievement criteria, the two sets of characteristics, other than the TEACHER and MEASURES x TEACHER characteristics, to account for a meaningful amount of criterion variance were the PERCEPT and MEASURES x PERCEPT characteristics. PERCEPT characteristics seem to explain covariation in ratings and achievement regardless of the TRF employed. PERCEPT characteristics also explain some of the variance in ratings independent of achievement (and vice-versa) regardless of the TRF employed.

Subsequent analyses found the minimum number of PERCEPT characteristics and MEASURES x PERCEPT characteristics necessary to account for a large percentage of the variance explained by the respective set. After entering the T, O, and T x O sets of characteristics each PERCEPT characteristic was entered in the analyses in a forward (stepwise)

manner. After entering the T through M x T x 0 sets of characteristics, each MEASURES x PERCEPT characteristic was also entered in the analyses in a forward (stepwise) manner. For each ratings/achievement criterion, the results were highly consistent. As shown in Table 8, students' perceptions of instructors' need for achievement (PACH) was always the first characteristic to enter the PERCEPT equations, was significantly related to the criterion, and generally accounted for 80% of the criterion variance explained by the set of PERCEPT characteristics. In addition, M x PACH was always the first characteristic to enter the M x P equations, was significantly related to the criterion, and generally accounted for 80% of the criterion variance explained by the set of M x P characteristics.

TABLE 8
 Forward (Stepwise) Repeated Measures MRAs to Find
 Best Subset of Predictors

	Source	SS	df	MS	F	p	Change in R^2	Proportion of Set
SULLIVAN & SKANES	PERCEPT (P)	144.089	4	36.023	35.282	.001	.071	1.000
	PACH	117.225	1	117.225	114.814	.001	.058	.817
	PU	14.778	1	14.778	14.474	.001	.007	.099
	PAFF	12.021	1	12.021	11.774	.001	.006	.085
	PD	0.064	1	0.064	0.063	.802	.000	.000
	Error	359.302	352	1.021				
	M x P	126.400	4	31.600	32.146	.001	.062	1.000
	M x PACH	103.002	1	103.002	104.783	.001	.051	.823
	M x PU	10.596	1	10.596	10.779	.001	.005	.081
	M x PAFF	12.267	1	12.267	12.479	.001	.006	.097
	M x PD	0.535	1	0.535	0.544	.461	.000	.000
	Error	346.143	352	0.983				
	HILDEBRAND & WILSON	PERCEPT (P)	508.749	4	127.187	41.169	.001	.062
PACH		429.996	1	429.996	140.705	.001	.052	.839
PD		38.879	1	38.879	12.722	.001	.005	.081
PAFF		20.232	1	20.232	6.620	.010	.002	.032
PU		19.642	1	19.642	6.427	.012	.002	.032
Error		1075.708	352	3.056				
M x P		528.417	4	132.104	44.964	.001	.057	1.000
M x PACH		402.334	1	402.334	136.941	.001	.049	.860
M x PAFF		39.862	1	39.862	13.568	.001	.005	.088
M x PU		25.325	1	25.325	8.620	.004	.003	.053
M x PD		3.912	1	3.912	1.332	.249	.000	.000
Error		1034.100	352	2.938				
POHLMANN		PERCEPT (P)	500.301	4	125.075	52.797	.001	.065
	PACH	395.617	1	395.617	166.997	.001	.051	.785
	PU	79.227	1	79.227	33.443	.001	.010	.154
	PAFF	22.345	1	22.345	9.432	.002	.003	.046
	PD	3.112	1	3.112	1.314	.253	.000	.000
	Error	833.989	352	2.369				
	M x P	461.462	4	115.366	51.503	.001	.060	1.000
	M x PACH	369.103	1	369.103	164.778	.001	.048	.800
	M x PU	69.088	1	69.088	30.843	.001	.009	.150
	M x PAFF	22.693	1	22.693	10.131	.002	.003	.050
	M x PD	0.578	1	0.578	0.258	.612	.000	.000
	Error	788.374	352	2.240				

DISCUSSION

The Student Opinion Construct

The present study failed to find evidence of a meaningful relationship between ratings and student personality characteristics independent of teacher characteristics. Neither the set of four OWN characteristics (which represented the four factors that accounted for most of the ACL variance) nor the set of 12 TEACHER by OWN interactions explained a significant amount of variance in any of the three rating criteria. This suggests that students form opinions about an instructor's teaching effectiveness independent of the students' personality characteristics. These findings are inconsistent with the conclusions of Follman (1975) and Feldman (1977) who suggested either a strong or moderate relationship existed between several student characteristics and ratings.

Based on prior evidence, it was further predicted that student characteristics more readily associated with the educational environment might relate to ratings. However, the set of OWN characteristics, which included a scale (ACH) related to an educational context, and the set of TEACHER by OWN characteristics never explained a significant amount of rating variance or accounted for much more than 1% of rating variance. Thus, no support was found for this prediction.

There are many possible explanations for these inconsistencies. Three appear most likely. First, most prior field research, where students were not randomly assigned to classes, used an undesirable unit of analysis--pooled students--to determine the relationship between student characteristics and ratings. Thus, contrary to the present study,

the relationship between ratings and student characteristics in these studies may not be independent of teacher effects on ratings. Second, the present use of a laboratory setting afforded a degree of experimental control typically unavailable in prior studies. While the laboratory setting increased the "internal validity" (Campbell & Stanley, 1963) of the study, it may have affected the "external validity" of the findings, making the results incomparable to prior field research. Third, this study is the first to use the ACL as a personality measure to examine the relationship between student characteristics and ratings. Unique properties of the ACL compared to other measures may also explain the inconsistencies.

Several researchers have been interested in the match, congruence, or similarity of students and teachers as it relates to student ratings. Researchers may investigate the perceived similarity between students and teachers or actual teacher-student similarity. The present study addressed researchers interested in the relationship between perceived similarity and ratings. But rather than compute the correlation between perceived similarity scores and ratings, TRF scores were regressed on the constituent variables of a perceived similarity score: student's own score and the student's ratings of the instructor's personality. As previously noted, the set of OWN characteristics did not account for any meaningful rating variance. However, for all three rating criteria, the set of PERCEPT characteristics accounted for at least 10% of the variance in ratings. How students perceive the personality characteristics of their instructors appears related to their teacher effectiveness ratings. Moreover, student perceptions of instructors' needs to achieve was the single best predictor of the set of PERCEPT characteristics. Interestingly,

this variable appears more closely related to an educational environment than the remaining PERCEPT characteristics.

At this point, researchers may want to reconsider the use of perceived similarity scores which require the measurement of OWN and PERCEPT scores. The present investigation demonstrated that a correlation between ratings and perceived similarity scores could be substantially explained by the correlation between ratings and PERCEPT scores alone. Unfortunately, the PERCEPT/ratings relationship is difficult to interpret; several explanations are possible. Students may rate their instructor's teaching effectiveness based on their impressions of instructor personality. According to this explanation, an instructor's ratings might improve if that instructor successfully altered student impressions of his/her personality. In support of this hypothesis, Ware and Williams (1975) found that lecturer expressiveness (i.e., enthusiasm, humor, charisma, etc.) affected teacher ratings. A second possible explanation is that students form impressions of instructor's personality based on their impressions of the instructor's teaching effectiveness. For example, students who highly rate an instructor might ascribe certain personal traits to that instructor while students who poorly rate that instructor might form different impressions of the instructor's personality. Yet another explanation is that students form impressions of both an instructor's personality and teaching effectiveness based on some other factor such as a student characteristic not identified in the present study. Further research is necessary to determine which of these alternatives is correct.

For researchers interested in the validity of the student opinion

construct, these data suggest that student personality characteristics do not account for variability in ratings which cannot be explained by teacher characteristics. These findings are evidence for the validity of ratings for researchers whose goal is a TRF which is not affected by characteristics presumed to be unrelated to teacher effectiveness. The relationship between PERCEPT characteristics and ratings found here may not be sufficient evidence to judge TRF validity. Because there are a variety of ways that this relationship may have occurred, care should be exercised in making any validity judgments.

Student Achievement

As might be expected, TEACHER characteristics explained a large percentage (30.8%) of the variance in student achievement. While several other sets of characteristics (i.e., OWN characteristics, TEACHER by OWN characteristics) were significantly related to achievement, they each accounted for less than 3% of the achievement variance. Only the set of TEACHER by PERCEPT interactions related meaningfully to student achievement. In general, this interaction suggests that the way teacher characteristics affect achievement is dependent on student perceptions of an instructor's personality. More specifically, one might consider only the teacher style by student perception of instructor deference (S x PD) factor, which accounted for 80% of the achievement variance explained by the set. This interaction suggests that the influence of an instructor's teaching style on achievement is mediated by the student's perceptions of his/her deference (i.e., degree of conscientiousness).

Thus, a particular teacher style may produce greater achievement only when the students perceive the teacher as having a certain degree of conscientiousness.

The significant set of TEACHER by PERCEPT interactions is consistent with a trait-treatment interaction approach to instructional effectiveness; TEACHER (treatment) and student (trait) characteristics interacted. However, the actual student trait involved here may be elusive. As suggested above, student differences in the PERCEPT characteristics may be due to a variety of factors including their expected achievement test score.

The Ratings/Achievement Relationship

For researchers interested in the ratings/achievement relationship, the factors of interest are those which: (a) explain variability in both ratings and achievement, and/or (b) explain variability in ratings independent of achievement (or vice-versa). For the set of OWN characteristics this amounts to examining both the OWN main effect and the OWN by MEASURES interaction. A significant OWN main effect suggests that variability in both ratings and achievement can be explained by student personality characteristics. A significant interaction suggests that student personality characteristics account for variability in ratings independent of achievement (or vice-versa). Finding a significant main effect or interaction is of interest when the correlation between ratings and achievement is computed across classmeans and students differ from class-to-class in mean student personality characteristics. When students are equivalent from class-to-class, the ratings/achievement correlation cannot be affected by student characteristics.

There was no evidence of a meaningful relationship between student personality characteristics and any of the three rating/achievement criteria. There were no significant OWN main effects or OWN by MEASURES interactions. In addition, the TEACHER by OWN and MEASURES by TEACHER by OWN interactions failed to reach significance.

There were significant PERCEPT main effects and PERCEPT by MEASURES interactions for all ratings/achievement criteria. The single best predictor of the set of PERCEPT characteristics was student perceptions of instructor's needs to achieve. This variable appears more closely related to an educational environment than the remaining PERCEPT characteristics. The single best predictor of the set of interactions was MEASURES by student perceptions of instructor's needs to achieve. These findings suggest that variability in ratings and achievement can be explained by PERCEPT characteristics. Also, variability in ratings independent of achievement (or vice-versa) can be explained by PERCEPT characteristics.

For researchers interested in the relationship between ratings and teacher-produced achievement these data suggest that student personality characteristics do not affect this relationship. When the correlation between ratings and achievement is computed across classmeans, the correlation will not be affected by student personality characteristics regardless of how students differ from class-to-class.

On the other hand, if students are not equivalent from class-to-class, the correlation between ratings and teacher-produced achievement may be affected by student perceptions of instructor personality. By itself, this may not be sufficient evidence to judge TRF validity because the process by which students form their impressions of instructor personality

is not known. In contrast, one might argue that these PERCEPT characteristics account for variability in ratings and achievement beyond that accounted for by teacher characteristics. And regardless of how students form judgments of instructor personality, this set of student characteristics may affect the correlation between ratings and teacher-produced achievement.

Cautions and Suggestions for Further Research

The cautions described in Leventhal et al. (1977) and Perry et al. (1978) concerning teacher effectiveness research conducted in laboratory settings apply to the present investigation. These include: (a) the experimental conditions differed from a real class; (b) the multiple-choice quiz may have been a poor psychometric device and probably measured only short-term learning; and (c) analyzing only the mean responses to TRF items ignores possible relationships between individual items or factors and teacher or student characteristics. In addition, there are other cautions unique to this study.

First, only linear relationships were examined in this study. Nonlinear relationships between student characteristics and the criterion measures were not examined. The addition of nonlinear components would have made the analyses unmanageably large and complicated. More importantly, there was not sufficient theoretical or empirical evidence available to suggest meaningful nonlinear relationships. Nevertheless, such relationships may exist.

Second, the ACL was used as a broadband measure of student personality. While it may measure a number of meaningful student personality

dimensions, by no means did it exhaust the entire spectrum. So, the finding of no relationship between student personality characteristics and either ratings or achievement should be understood in this light.

Third, students participated in the study in two hour sessions. They completed the ACL for the instructor immediately after completing the TRF and quiz. Responses to the teacher ACL might have been different if the students were less tired or had more of an opportunity to reflect on the personality characteristics of the videotaped lecturer.

The results of this study should be replicated in a non-laboratory setting. Several sections of a multisection course, taught by different instructors, and employing a common final exam would be needed to confirm the present findings under more natural classroom conditions. In addition, it would be of interest to determine the degree to which students differ from class-to-class on OWN and PERCEPT characteristics. If students are not different between classes then neither OWN nor PERCEPT characteristics can affect the ratings/achievement relationship computed across class section means. Finally, the processes that affect student PERCEPTS need to be probed. This might be accomplished in several ways. Generally, one might question students about how they form opinions of an instructor's personality. More specifically, it may be of interest to determine whether PERCEPT judgments are based on student evaluations of teaching. Here, one might compare the PERCEPT scores assigned to teachers observed teaching or engaged in some non-instructional activity. Also, the effects of testing on PERCEPT scores might be probed. Here, one might compare quiz and no quiz students.

REFERENCE NOTES

1. Whitely, S.E. On the invalidity of construct validity for student ratings. Paper presented at the annual meeting of American Education Research Association, April, 1975.
2. Leventhal, L., Abrami, P.C., and Perry, R.P. Bogus evidence for the validity of student ratings. Paper presented at the annual meeting of the American Psychological Association, San Francisco, August, 1977.
3. Abrami, P.C., Leventhal, L., and Perry, R.P. The generalizability of student ratings of instruction. Paper presented at the annual meeting of the American Educational Research Association, New York, April, 1977.
4. McKeachie, W.J. Personal communication, January 9, 1978.
5. Perry, R.P., Leventhal, L., and Abrami, P.C. Standardization study. Manuscript in preparation, University of Manitoba, 1978.

REFERENCES

- American Psychological Association. Standards for educational and psychological tests. Washington, D.C.: Author, 1974.
- Bausell, R., & Magoon, J. Instructional methods and college student ratings of courses and instructors. Journal of Experimental Education, 1972, 40, 29-33.
- Berliner, D.C., & Cahen, L.S. Trait-treatment interaction and learning. Review of Research in Education, 1973, 1, 58-94.
- Blass, T. Measurement of objectivity-subjectivity: Effects of tolerance for imbalance and grades on evaluations of teachers. Psychological Reports, 1974, 34, 1199-1213.
- Blass, T. Locus of control, grades, and course evaluations. Personality and Social Psychology Bulletin, 1976, 2, 79-81.
- Buros, O.K. (Ed.). Personality tests and reviews. Highland Park, N.J.: Gryphon Press, 1970.
- Buros, O.K. (Ed.). The seventh mental measurements yearbook. Highland Park, N.J.: Gryphon Press, 1972.
- Campbell, D.T., & Fiske, D.W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.
- Campbell, D.T., & Stanley, J.C. Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.) Handbook of research on teaching. Chicago: Rand McNally, 1963.
- Cooley, W., & Lohnes, P. Multivariate data analysis. New York: Wiley, 1971.

- Costin, F., & Grush, J.E. Personality correlates of teacher-student behavior in the college classroom. Journal of Educational Psychology, 1973, 65, 35-44.
- Cronbach, L.J., & Snow, R.F. Aptitude and instructional methods. New York: Irvington Press, 1977.
- Crittenden, K.S., & Norr, J.L. Student values and teacher evaluation: A problem in person perception. Sociometry, 1973, 36, 143-151.
- Davison, D.C. Perception of instructor in relation to self and evaluation of instructor's performance. Perceptual and Motor Skills, 1973, 36, 533-534.
- Domino, G. Interactive effects of achievement orientation and teaching style on academic achievement. Journal of Educational Psychology, 1971, 62, 427-431.
- Doty, B.A. Teaching method effectiveness in relation to certain student characteristics. Journal of Educational Research, 1967, 60, 363-365.
- Doyle, K.O., & Whitely, S.E. Student ratings as criteria for effective teaching. American Educational Research Journal, 1974, 11, 259-274.
- Feldhusen, J.S., & Starks, D.D. Bias in college students' ratings of instructors. College Student Survey, 1970, 4, 6-9.
- Feldman, K.A. Consistency and variability among college students in rating their teachers and courses: A review and analysis. Research in Higher Education, 1977, 6, 223-274.
- Follman, J. Student ratings of faculty teaching effectiveness: Rater or ratee characteristics? Research in Higher Education, 1975, 3, 155-167.
- Freehill, M. Authoritarian bias and evaluation of college experiences. Improving College and University Teaching, 1967, 15, 18-19.

- Fulcher, D.G., and Anerson, W.T. Interpersonal dissimilarity and teaching effectiveness: A relational analysis. Journal of Educational Research, 1974, 68, 19-25.
- Gabriel, R.M., & Hopkins, K.D. Relative merits of MANOVA, repeated measures ANOVA, and univariates ANOVAs for research utilizing multiple criterion measures. Journal of Special Education, 1974, 8, 377-389.
- Gage, N. The appraisal of college teaching. Journal of Higher Education, 1961, 32, 17-22.
- Good, K.C., & Good, L.R. Assumed attitude similarity and instructor evaluation. Journal of Social Psychology, 1973, 91, 285-290.
- Good, K.C., & Good, L.R. Attitude similarity and attraction to an instructor. Psychological Reports, 1973, 33, 335-337.
- Gough, H.G., & Heilbrun, A.B., Jr. Adjective check list. Palo Alto, Calif.: Consulting Psychologists Press, 1967.
- Granzin, K.L., & Painter, J.J. A new explanation for students' course evaluation tendencies. American Educational Research Journal, 1973, 10, 115-124.
- Grush, J.E., Clore, G.L., & Costin, F. Dissimilarity and attraction: When difference makes a difference. Journal of Personality and Social Psychology, 1975, 32, 783-789.
- Grush, J.E., & Costin, F. Student or consumer of the teaching process: Variables important for teacher effectiveness. American Educational Research Journal, 1975, 12, 55-66.
- Hall, D. The effect of teacher-student congruence upon student learning in college classes. Journal of Educational Psychology, 1970, 61, 205-213.

- Harris, R.J. A primer of multivariate statistics. New York: Academic Press, 1975.
- Haslett, B.J. Attitudes toward teachers as a function of student academic self-concept. Research in Higher Education, 1976, 4, 41-58.
- Hays, W.L. Statistics for the social sciences (2nd Ed.). New York: Holt, Rinehart, and Winston, 1973.
- Hildebrand, M., & Wilson, R. Effective university teaching and its evaluation. Washington, D.C.: U.S. Department of Health, Education, and Welfare, Office of Education, 1970.
- Isaacson, R.L., McKeachie, W.J., Milholland, J.E., Lin, Y.G., Hofeller, M., Baerwaldt, J.W., & Zinn, K.L. Dimensions of student evaluations of teaching. Journal of Educational Psychology, 1964, 55, 344-351.
- Kerlinger, F.N., & Pedhazur, E.J. Multiple regression in behavioral research. New York: Holt, Rinehart, and Winston, 1973.
- Kovacs, R., & Kapel, D.E. Personality correlates of faculty and course evaluations. Research in Higher Education, 1976, 5, 335-344.
- Levenson, H., & LeUnes, A. Student evaluation of an instructor: Effects of attitude similarity. Psychological Reports, 1974, 34, 1074.
- Leventhal, L. Teacher ratings forms: Critique and reformulation of previous validation designs. Canadian Psychology Review, 1975, 16, 269-276.
- Leventhal, L., Abrami, P.C., & Perry, R.P. Do teacher rating forms reveal as much about students as about teachers? Journal of Educational Psychology, 1976, 68, 441-445.
- Leventhal, L., Abrami, P.C., Perry, R.P., & Breen, L.J. Section selection in multi-section courses: Implications for the validation and

- use of teacher rating forms. Educational and Psychological Measurement, 1975, 35, 885-895.
- Levinthal, C., Lansky, L., & Andrews, O. Student evaluations of teacher behaviors as estimations of real-ideal discrepancies: A critique of teacher rating methods. Journal of Educational Psychology, 1971, 62, 104-109.
- Maney, A.C. The authoritarianism dimension in student evaluation of faculty. Journal of Educational Sociology, 1959, 32, 226-231.
- Menges, R.J. Student-instructor cognitive compatibility in the large lecture class. Journal of Personality, 1969, 37, 444-458.
- Morstain, B. Relationship of student and instructor educational orientations with course ratings. Journal of Educational Psychology, 1977, 69, 338-396.
- Null, E.J., & Walter, J.E. Values of students and their ratings of a university professor. College Student Journal, 1972, 6, 46-51.
- Page, M.M. Correlation between student estimated grades and actual grades: A reply to Blass. Personality and Social Psychology Bulletin, 1976, 2, 82-83.
- Page, M.M., & Roy, R.E. Internal-external control and independence of judgment in course evaluations among college students. Personality and Social Psychology Bulletin, 1975, 1, 509-512.
- Painter, J., & Granzin, K. Consistency theory as an explanation of students' course evaluation tendencies. Journal of Experimental Education, 1972, 41, 78-81.
- Parent, J., Forward, J., Canter, R., & Mohling, J. Interactive effects of teaching strategy and personal locus of control on student performance and satisfaction. Journal of Educational Psychology, 1975, 67, 764-769.

- Pedhazur, E.J. Coding subjects in repeated measures designs. Psychological Bulletin, 1977, 84, 298-305.
- Perry, R.P., Abrami, P.C., & Leventhal, L. Educational seduction: The effect of instructor expressiveness and lecture content on student ratings and achievement. Journal of Educational Psychology, 1978, in press.
- Pohlmann, J.T. A multivariate analysis of selected class characteristics and student ratings of instruction. Multivariate Behavioral Research, 1975, 10, 81-92.
- Rees, R. Dimensions of students' points of view in rating college teachers. Journal of Educational Psychology, 1969, 60, 476-482.
- Rezler, A.G. The influence of needs upon the student's perception of his instructor. Journal of Educational Research, 1965, 58, 282-286.
- Rodin, M., & Rodin, B. Student evaluations of teachers. Science, 1972, 177, 1164-1166.
- Sherman, B.R., & Blackburn, R.T. Personal characteristics and teaching effectiveness of college faculty. Journal of Educational Psychology, 1975, 67, 124-131.
- Sullivan, A.M., & Skanes, G.R. Validity of student evaluation of teaching and the characteristics of successful instructors. Journal of Educational Psychology, 1974, 66, 584-590.
- Taguiri, R. Person perception. In G. Lindzey and E. Aronson (Eds.), The handbook of social psychology (2nd Ed.), Vol. 3. Reading, Mass.: Addison-Wesley, 1969.
- Tetenbaum, T.J. The role of student needs and teacher orientations in student ratings of teachers. American Educational Research Journal,

- 1975, 12, 417-429.
- Tobias, S. Achievement treatment interactions. Review of Educational Research, 1976, 46, 61-74.
- Tobias, S., & Hanlon, R. Attitudes towards instructors, social desirability, and behavioral intentions. Journal of Educational Psychology, 1975, 67, 405-408.
- Tuckman, B.W., & Orefice, D.S. Personality structure, instructional outcomes, and instructional preferences. Interchange, 1973, 4, 43-48.
- Ware, J.E., & Williams, R.G. The Doctor Fox Effect: A study of lecturer effectiveness and ratings of instruction. Journal of Medical Education, 1975, 50, 149-156.
- White, W.F., & Wash, J.A., Jr. Perception of teacher effectiveness as a function of the students' need for social approval. Perceptual and Motor Skills, 1966, 23, 711-717.
- Yonge, G., & Sassenrath, J. Student personality correlates of teacher ratings. Journal of Educational Psychology, 1968, 59, 44-52.

APPENDIX

Written Materials Employed in the Study

INSTRUCTIONS

Because of the growing interest in improving university instruction, valid methods for measuring faculty teaching performance are in great demand. Since the student is the consumer of the teaching process, the student seems to be the best person to evaluate it. Student evaluations of courses are therefore increasingly important at the University of Manitoba and other universities throughout the world.

Unfortunately, it is very difficult to determine whether student ratings really are valid measures of teaching effectiveness. There may be some unknown, uncontrolled processes which influence the outcome of student evaluations and make the results less useful. One way to solve this problem is to conduct laboratory experiments where the factors that might influence the outcome of student evaluations are controlled.

Therefore, we are conducting a series of controlled experiments to determine the validity of student ratings of university instructors. In particular, the present study is concerned with the reactions of different types of students to university instructors. Your first task today will be to describe yourself on an adjective check list so that we can learn something about you. You will then be shown an unedited videotape of a lecture delivered by an instructor who has volunteered to participate in this experiment. Following the videotape, you will be asked to give your impressions of the instructor's teaching performance. In addition, you will be given a short quiz to determine how much you recall of the lecture. (Since you may want to take notes on the lecture, you have been provided with a pencil and paper.) Finally, we will ask you to describe your personal impressions of the instructor on an adjective check list.

Please note that it is important that other people participating in this experiment after you have no preconceived notions about it. Therefore, please do not discuss this experiment with anyone after you leave. If you have any questions about what to do during the experiment, please raise your hand; do not shout questions out. A more complete description of this study will be given at the end of today's session.

ADJECTIVE CHECK LIST TO DESCRIBE YOURSELF

NOTE: For this adjective check list, use IBM answer sheets A and B.

The following two pages contain a list of adjectives. Please read them quickly. If you consider an adjective to be self-descriptive, fill in alternative "T" (true) on the appropriate IBM answer sheet. If the adjective does not describe you, leave the alternative blank. Do not worry about duplications, contradictions, and so forth. Work quickly and do not spent too much time on any one adjective. Try to be frank, and select those adjectives which describe you as you really are, not as you would like to be.

NOTE: For the adjectives appearing below, use IBM answer sheet "A".

A1	absent-minded	A51	cowardly	A101	gloomy
A2	active	A52	cruel	A102	good-looking
A3	adaptable	A53	curious	A103	good-natured
A4	adventurous	A54	cynical	A104	greedy
A5	affected	A55	daring	A105	handsome
A6	affectionate	A56	deceitful	A106	hard-headed
A7	aggressive	A57	defensive	A107	hard-hearted
A8	alert	A58	deliberate	A108	hasty
A9	aloof	A59	demanding	A109	headstrong
A10	ambitious	A60	dependable	A110	healthy
A11	anxious	A61	dependent	A111	helpful
A12	apathetic	A62	despondent	A112	high-strung
A13	appreciative	A63	determined	A113	honest
A14	argumentative	A64	dignified	A114	hostile
A15	arrogant	A65	discreet	A115	humorous
A16	artistic	A66	disorderly	A116	hurried
A17	assertive	A67	dissatisfied	A117	idealistic
A18	attractive	A68	distractible	A118	imaginative
A19	autocratic	A69	distrustful	A119	immature
A20	awkward	A70	dominant	A120	impatient
A21	bitter	A71	dreamy	A121	impulsive
A22	blustery	A72	dull	A122	independent
A23	boastful	A73	easy going	A123	indifferent
A24	bossy	A74	effeminate	A124	individualistic
A25	calm	A75	efficient	A125	industrious
A26	capable	A76	egotistical	A126	infantile
A27	careless	A77	emotional	A127	informal
A28	cautious	A78	energetic	A128	ingenious
A29	changeable	A79	enterprising	A129	inhibited
A30	charming	A80	enthusiastic	A130	initiative
A31	cheerful	A81	evasive	A131	insightful
A32	civilized	A82	excitable	A132	intelligent
A33	clear-thinking	A83	fair-minded	A133	interests narrow
A34	clever	A84	fault-finding	A134	interests wide
A35	coarse	A85	fearful	A135	intolerant
A36	cold	A86	feminine	A136	inventive
A37	commonplace	A87	fickle	A137	irresponsible
A38	complaining	A88	flirtatious	A138	irritable
A39	complicated	A89	foolish	A139	jolly
A40	conceited	A90	forceful	A140	kind
A41	confident	A91	foresighted	A141	lazy
A42	confused	A92	forgetful	A142	leisurely
A43	conscientious	A93	forgiving	A143	logical
A44	conservative	A94	formal	A144	loud
A45	considerate	A95	frank	A145	loyal
A46	contented	A96	friendly	A146	mannerly
A47	conventional	A97	frivolous	A147	masculine
A48	cool	A98	fussy	A148	mature
A49	cooperative	A99	generous	A149	meek
A50	courageous	A100	gentle	A150	methodical

REMEMBER: SELF-DESCRIPTIVE = "T" (true)
NOT SELF-DESCRIPTIVE = BLANK

NOTE: For the adjectives appearing below, use IBM answer sheet "B".

B1	mild	B51	reliable	B101	strong
B2	mischievous	B52	resentful	B102	stubborn
B3	moderate	B53	reserved	B103	submissive
B4	modest	B54	resourceful	B104	suggestible
B5	moody	B55	responsible	B105	sulky
B6	nagging	B56	restless	B106	superstitious
B7	natural	B57	retiring	B107	suspicious
B8	nervous	B58	rigid	B108	sympathetic
B9	noise	B59	robust	B109	tactful
B10	obliging	B60	rude	B110	tactless
B11	obnoxious	B61	sarcastic	B111	talkative
B12	opinionated	B62	self-centered	B112	temperamental
B13	opportunistic	B63	self-confident	B113	tense
B14	optimistic	B64	self-controlled	B114	thankless
B15	organized	B65	self-denying	B115	thorough
B16	original	B66	self-pitying	B116	thoughtful
B17	outgoing	B67	self-punishing	B117	thrifty
B18	outspoken	B68	self-seeking	B118	timid
B19	painstaking	B69	selfish	B119	tolerant
B20	patient	B70	sensitive	B120	touchy
B21	peaceable	B71	sentimental	B121	tough
B22	peculiar	B72	serious	B122	trusting
B23	persevering	B73	severe	B123	unaffected
B24	persistent	B74	sexy	B124	unambitious
B25	pessimistic	B75	shallow	B125	unassuming
B26	planful	B76	sharp-witted	B126	unconventional
B27	pleasant	B77	shiftless	B127	undependable
B28	pleasure-seeking	B78	show-off	B128	understanding
B29	poised	B79	shrewd	B129	unemotional
B30	polished	B80	shy	B130	unexcitable
B31	practical	B81	silent	B131	unfriendly
B32	praising	B82	simple	B132	uninhibited
B33	precise	B83	sincere	B133	unintelligent
B34	prejudiced	B84	slipshod	B134	unkind
B35	preoccupied	B85	slow	B135	unrealistic
B36	progressive	B86	sly	B136	unscrupulous
B37	prudish	B87	smug	B137	unselfish
B38	quarrelsome	B88	snobbish	B138	unstable
B39	queer	B89	sociable	B139	vindictive
B40	quick	B90	soft-hearted	B140	versatile
B41	quiet	B91	sophisticated	B141	warm
B42	quitting	B92	spendthrift	B142	wary
B43	rational	B93	spineless	B143	weak
B44	rattlebrained	B94	spontaneous	B144	whiny
B45	realistic	B95	spunky	B145	wholesome
B46	reasonable	B96	stable	B146	wise
B47	rebellious	B97	steady	B147	withdrawn
B48	reckless	B98	stern	B148	witty
B49	reflective	B99	stingy	B149	worrying
B50	relaxed	B100	stolid	B150	zany

REMEMBER: SELF-DESCRIPTIVE = "T" (true)
NOT SELF-DESCRIPTIVE = BLANK

TEACHER EVALUATION

NOTE: For the teacher evaluation, use items numbered 1 - 44 on IBM answer sheet "C".

The following questions are intended to evaluate your impressions of the instructor in the videotape. Base your responses on the instructor's performance you have just observed and what you expect your reactions would be if you were to take a course from the instructor. Some of the questions you may find difficult to answer but nevertheless, take a best guess. PLEASE DO NOT LEAVE ANY ANSWERS BLANK. Please mark your responses on the IBM answer sheet provided. Do not mark this form. Also, please enter only your name and not your student number on the IBM answer sheet.

- C1. How would you rate the instructor in general (all-around) teaching ability?
- (a) poor and inadequate instructor
 - (b) an adequate, but not stimulating instructor
 - (c) a good instructor
 - (d) a very good instructor
 - (e) an outstanding and stimulating instructor

Each of the following statements describes a basic component of teaching. Give the instructor an overall rating for each component, reserving the highest scores for unusually effective performance. Use the following scale to determine your rating:

- (a) well below average
- (b) below average
- (c) average
- (d) above average
- (e) well above average

- C2. Discuss points of view other than his own.
- C3. Explains clearly.
- C4. Encourage class discussion.
- C5. Has a genuine interest in students.
- C6. Is a dynamic and energetic person.
- C7. Has an interesting style of presentation.
- C8. Contrasts implications of various theories.
- C9. Invites students to share their knowledge and experiences.
- C10. Is friendly toward students.
- C11. Is well prepared.
- C12. Discusses recent developments in the field.
- C13. Presents origins of ideas and concepts.
- C14. Gives lectures that are easy to outline.
- C15. Relates to students as individuals.
- C16. Seems to enjoy teaching.
- C17. Gives references for more interesting and involved points.
- C18. Is enthusiastic about his subject.
- C19. Seems to have self-confidence.
- C20. Recognizes and greets students out of class.
- C21. Clarifies thinking by identifying reasons for questions.
- C22. Invites criticism of his own ideas.
- C23. Is careful and precise in answering questions.
- C24. Summarizes major points.
- C25. Is accessible to students out of class.
- C26. Knows if class is understanding him or not.

Below you will find an additional series of statements about the lecturer in the videotape you just saw. Please read each statement carefully, and respond to each using the following scale:

- (a) Improvement definitely needed
- (b) Weak performance
- (c) Good performance
- (d) Very good performance
- (e) Exceptional performance

THE LECTURER:

- C27. Spoke understandably.
- C28. Knew if students understood him.
- C29. Showed an interest in students.
- C30. Increased your appreciation for the subjects.
- C31. Gave several examples to explain complex ideas.
- C32. Knew his subject matter.

Below you will find another series of statements about the lecturer in the videotape you just saw. Please read each statement carefully, and respond to each using the following scale:

- (a) I strongly disagree with the statement.
- (b) I disagree with the statement.
- (c) Neutral, I neither agree nor disagree.
- (d) I agree with the statement.
- (e) I strongly agree with the statement.

THE LECTURER:

- C33. Stressed important material.
- C34. Was an effective lecturer.
- C35. Has a good sense of humour.
- C36. Organized and presented subject matter well.
- C37. Inspired confidence in his knowledge of the subject.
- C38. Broadened my interest in the subject.
- C39. Explained the subject clearly.
- C40. Increased my knowledge of the subject.
- C41. Stimulated my thinking.
- C42. Was enthusiastic about the subject.
- C43. Made learning enjoyable.
- C44. Did not often leave me confused.

QUIZ

NOTE: For the quiz, use items numbered 45 - 75 on IBM answer sheet "C".

DIRECTIONS: This is a quiz over the lecture on sex roles. Please answer the questions to the best of your ability. No one is expected to get all the answers correct. If you are in doubt about the answer to a question, then guess.

Please place your answers on the computed-scored answer sheet which has been provided. Please do not mark in the test booklet.

All responses must be made using the pencil provided. Please choose the one best answer for each item.

- C45. "Sex role identity" is
- the same as sexual preference
 - one component of sexual identity
 - the same as gender identity
 - bipolar
- C46. Traditional conceptions of sex role identity have been _____, in contrast to the notion of androgyny, which conceptualizes sex-role identity as _____. (Choose the alternative that fills in the blanks in the correct order.)
- masculine; feminine
 - feminine; masculine
 - one dimensional; two dimensional
 - two dimensional; one dimensional
- C47. An androgynous person would be
- high on positive masculine characteristics and low on feminine characteristics
 - low on positive masculine characteristics and high on feminine characteristics
 - high on both positive masculine and feminine characteristics
 - low on both masculine and feminine characteristics
- C48. Older evidence suggests that strongly sex-typed males, compared to less "masculine" males
- are better adjusted throughout the life span
 - are less well adjusted throughout the life span
 - are better adjusted in adolescence but not later
 - are less well adjusted in adolescence but not later
- C49. The major advantage of being androgynous is, in general terms:
- independence
 - nurturance and intimacy
 - financial
 - flexibility
- C50. In an experiment in which subjects had to perform cross-sex behaviors
- androgynous people felt badly about themselves
 - strongly sex-typed people felt badly about themselves
 - androgynous peoples expressed anger at being forced to do the behaviors
 - sex-typed people were more successful at the task.
- C51. The group(s) that did best in the experiments involving judgements of humour were
- masculine and feminine subjects
 - androgynous and feminine subjects
 - masculine and androgynous subjects
 - masculine subjects

- C52. In the experiments in which subjects were asked to converse with another person, the group(s) that did best were
- masculine and feminine subjects
 - androgynous and feminine subjects
 - masculine and androgynous subjects
 - masculine subjects
- C53. An androgynous model of sex-roles would move the traditional male sex-role toward a greater emphasis on
- intimacy and self-esteem
 - independence and self-esteem
 - intimacy and experiencing of emotions
 - self-esteem and experiencing of emotions
- C54. The interpersonal behavior circle describes behavior as falling along what two dimensions?
- acceptance-rejection
 - dominance-femininity
 - dominance-acceptance
 - dominance-androgyny
- C55. Most traditional religions have treated women
- as inferior
 - with respect
 - as man's equal
 - as superior
- C56. The study in which people were asked to judge the severity of disturbance indicated in brief written descriptions of therapy illustrated the price one pays for
- being too strongly sex-typed
 - deviating from traditional sex-roles
 - being androgynous
 - being overly feminine
- C57. A study in which ratings were done of identical articles, on which only the sex of the author was changed, showed
- "feminine appropriate" topics authored by female authors were rated higher
 - having a male author resulted in higher ratings for both "female" appropriate and "male appropriate" topics
 - having a male author resulted in higher rating for only "male appropriate" topics
 - none of a, b, or c is a correct statement

- C58. In the study of choices among sex-typed behaviors, the most money was sacrificed by
- strongly sex-typed people
 - people with virtually no sex-role identity
 - androgynous people
 - masculine typed people
- C59. The lecturer stressed four characteristics of the effective person:
- sex-role appropriate behavior, sexual preference, high self-esteem, openness to crying behaviors
 - intimacy, sex-role appropriate behavior, openness to emotions, humbleness
 - high self-esteem, intimacy, openness to emotions, independence
 - moderate self-esteem, sex-role appropriate behavior, intimacy, independence
- C60. On the interpersonal behavior circle, the traditional masculine sex-role would be best defined by the words
- friendly-dominant
 - hostile dominant
 - active-dominant
 - passive-dominant
- C61. The term "gender identity" refers to
- one's attitude toward people who engage in sex-role inappropriate behaviors
 - one's attitude toward people of the same sex, with regard to sexual behavior
 - the person's attitude toward his or her own body
 - the same thing as "sex-role identity"
- C62. An androgynous model of sex roles would move the traditional female sex role toward a greater emphasis on
- intimacy and self-esteem
 - independence and self-esteem
 - intimacy and experiencing of emotions
 - self-esteem and experiencing of emotions
- C63. The experiments dealing with judgements of humour were measuring
- ability to sense nuances of mood in others
 - complexity of understanding of humorous messages
 - the continuous or discontinuous nature of bipolar dimensions
 - independence of judgement
- C64. The lecturer pointed out that effeminate behaviors in males
- are innate in some people
 - are not natural behaviors
 - are only subtly different from masculine behaviors
 - are correlated with homosexuality

- C65. In one study, professionals made judgements on what emotionally healthy people are like. It was found that
- their description of a "healthy person" fit their descriptions of a "healthy man" and "healthy woman" equally well
 - their description of a "healthy person" was more like the "healthy man" than the "healthy woman"
 - their descriptions of a "healthy person" was more like the "healthy woman" than the "healthy man"
 - none of a, b, or c, is a correct statement of the results of the study
- C66. When the lecturer referred to masculinity and femininity as orthogonal components of androgyny, he meant that
- they are independent components
 - they are on a bipolar scale
 - they are continuous variables, in the sense that there is no clear dividing line where one stops and the other starts
 - they are incompatible with each other
- C67. Strongly sex-typed feminine women tend to
- be more loving than androgynous women
 - have lower self-esteem
 - be more controlling
 - avoid talking with other people
- C68. In Bem's study of the performance of sex-typed behaviors, the most money was offered for choosing
- an equal balance among the behaviors offered
 - sex-role appropriate behaviors
 - cross-sex behaviors
 - feminine-appropriate behaviors
 - male-appropriate behaviors
- C69. The lecturer said that traditional sex-roles
- are natural consequences of inborn sex-linked characteristics
 - although not necessarily natural, are the most adaptive arrangement for human society
 - should be abandoned because the evidence clearly indicates that no characteristics should be gender-linked
 - needlessly limit people's lives in some ways
- C70. The lecturer suggested that sexism
- probably exists in everyone to some extent
 - can be eliminated by receiving training as a mental health professional
 - serves some valid functions in society as it is currently structured
 - both a and c are correct

- C71. The group that would be most likely to be high on anxiety is
- strongly sex-typed feminine women
 - strongly sex-typed masculine men
 - strongly androgynous middle-aged men
 - strongly androgynous middle-aged women
- C72. Whether a person is masculine, feminine or androgynous is related to his/her
- sexual preference
 - sex-role identity
 - gender identity
 - place on the bipolar continuum
- C73. The lecturer stressed two main perspectives for the listener to take in understanding the significance of androgyny:
- one's view of oneself and of society, with regard to gender identity
 - one's view of others and of society, with regard to sexual preference
 - one's view of oneself and of others, with regard to sex-roles
 - one's view of oneself and of others, with regard to sexual preferences
- C74. The lecturer suggested that many behavioral differences that define effeminate and masculine actions are
- the result of differences in bone structure
 - subtle and learned
 - irrelevant to other's view of what the person has to say
 - natural

Please answer the following question on the IBM sheet:

- C75. Have you had this material (lecture) in class?
- Yes
 - No

ADJECTIVE CHECK LIST TO DESCRIBE THE INSTRUCTOR

NOTE: For this adjective check list, use IBM answer sheets D and E.

The following two pages contain a list of adjectives. Please read them quickly. If you consider an adjective to be descriptive of the instructor in the videotape, fill in alternative 'T' (true) on the appropriate IBM answer sheet. If the adjective does not describe the instructor, leave the alternative blank. Do not worry about duplications, contradictions, and so forth. Work quickly and do not spend too much time on any one adjective. Try to be frank, and select those adjectives which describe the instructor as he really is, not as you would like him to be.

NOTE: For the adjectives appearing below, use IBM answer sheet "D".

D1	absent-minded	D51	cowardly	D101	gloomy
D2	active	D52	cruel	D102	good-looking
D3	adaptable	D53	curious	D103	good-natured
D4	adventurous	D54	cynical	D104	greedy
D5	affected	D55	daring	D105	handsome
D6	affectionate	D56	deceitful	D106	hard-headed
D7	aggressive	D57	defensive	D107	hard-hearted
D8	alert	D58	deliberate	D108	hasty
D9	aloof	D59	demanding	D109	headstrong
D10	ambitious	D60	dependable	D110	healthy
D11	anxious	D61	dependent	D111	helpful
D12	apathetic	D62	despondent	D112	high-strung
D13	appreciative	D63	determined	D113	honest
D14	argumentative	D64	dignified	D114	hostile
D15	arrogant	D65	discreet	D115	humorous
D16	artistic	D66	disorderly	D116	hurried
D17	assertive	D67	dissatisfied	D117	idealistic
D18	attractive	D68	distractible	D118	imaginative
D19	autocratic	D69	distrustful	D119	immature
D20	awkward	D70	dominant	D120	impatient
D21	bitter	D71	dreamy	D121	impulsive
D22	blustery	D72	dull	D122	independent
D23	boastful	D73	easy going	D123	indifferent
D24	bossy	D74	effeminate	D124	individualistic
D25	calm	D75	efficient	D125	industrious
D26	capable	D76	egotistical	D126	infantile
D27	careless	D77	emotional	D127	informal
D28	cautious	D78	energetic	D128	ingenious
D29	changeable	D79	enterprising	D129	inhibited
D30	charming	D80	enthusiastic	D130	initiative
D31	cheerful	D81	evasive	D131	insightful
D32	civilized	D82	excitable	D132	intelligent
D33	clear-thinking	D83	fair-minded	D133	interests narrow
D34	clever	D84	fault-finding	D134	interests wide
D35	coarse	D85	fearful	D135	intolerant
D36	cold	D86	feminine	D136	inventive
D37	commonplace	D87	fickle	D137	irresponsible
D38	complaining	D88	flirtatious	D138	irritable
D39	complicated	D89	foolish	D139	jolly
D40	conceited	D90	forceful	D140	kind
D41	confident	D91	foresighted	D141	lazy
D42	confused	D92	forgetful	D142	leisurely
D43	conscientious	D93	forgiving	D143	logical
D44	conservative	D94	formal	D144	loud
D45	considerate	D95	frank	D145	loyal
D46	contented	D96	friendly	D146	mannerly
D47	conventional	D97	frivolous	D147	masculine
D48	cool	D98	fussy	D148	mature
D49	cooperative	D99	generous	D149	meek
D50	courageous	D100	gentle	D150	methodical

REMEMBER: DESCRIPTIVE OF THE INSTRUCTOR = "T" (true)
 NOT DESCRIPTIVE OF THE INSTRUCTOR = BLANK

NOTE: For the adjectives appearing below, use IBM answer sheet "E".

E1	mild	E51	reliable	E101	strong
E2	mischievous	E52	resentful	E102	stubborn
E3	moderate	E53	reserved	E103	submissive
E4	modest	E54	resourceful	E104	suggestible
E5	moody	E55	responsible	E105	sulky
E6	nagging	E56	restless	E106	superstitious
E7	natural	E57	retiring	E107	suspicious
E8	nervous	E58	rigid	E108	sympathetic
E9	noisy	E59	robust	E109	tactful
E10	obliging	E60	rude	E110	tactless
E11	obnoxious	E61	sarcastic	E111	talkative
E12	opinionated	E62	self-centered	E112	temperamental
E13	opportunistic	E63	self-confident	E113	tense
E14	optimistic	E64	self-controlled	E114	thankless
E15	organized	E65	self-denying	E115	thorough
E16	original	E66	self-pitying	E116	thoughtful
E17	outgoing	E67	self-punishing	E117	thrifty
E18	outspoken	E68	self-seeking	E118	timid
E19	painstaking	E69	selfish	E119	tolerant
E20	patient	E70	sensitive	E120	touchy
E21	peaceable	E71	sentimental	E121	tough
E22	peculiar	E72	serious	E122	trusting
E23	persevering	E73	severe	E123	unaffected
E24	persistent	E74	sexy	E124	unambitious
E25	pessimistic	E75	shallow	E125	unassuming
E26	planful	E76	sharp-witted	E126	unconventional
E27	pleasant	E77	shiftless	E127	undependable
E28	pleasure-seeking	E78	show-off	E128	understanding
E29	poised	E79	shrewd	E129	unemotional
E30	polished	E80	shy	E130	unexcitable
E31	practical	E81	silent	E131	unfriendly
E32	praising	E82	simple	E132	uninhibited
E33	precise	E83	sincere	E183	unintelligent
E34	prejudiced	E84	slipshod	E184	unkind
E35	preoccupied	E85	slow	E135	unrealistic
E36	progressive	E86	sly	E136	unscrupulous
E37	prudish	E87	smug	E137	unselfish
E38	quarrelsome	E88	snobbish	E138	unstable
E39	queer	E89	sociable	E139	vindictive
E40	quick	E90	soft-hearted	E140	versatile
E41	quiet	E91	sophisticated	E141	warm
E42	quitting	E92	spendthrift	E142	wary
E43	rational	E93	spineless	E143	weak
E44	rattlebrained	E94	spontaneous	E144	whiny
E45	realistic	E95	spunky	E145	wholesome
E46	reasonable	E96	stable	E146	wise
E47	rebellious	E97	steady	E147	withdrawn
E48	reckless	E98	stern	E148	witty
E49	reflective	E99	stingy	E149	worrying
E50	relaxed	E100	stolid	E150	zany

REMEMBER: DESCRIPTIVE OF THE INSTRUCTOR = "T" (true)
 NOT DESCRIPTIVE OF THE INSTRUCTOR = BLANK