AN EXPLORATORY STUDY OF THE RELATIONSHIP OF STUDY HABITS, EXPECTED GRADE AND INSTRUCTOR EFFECTIVENESS RATINGS TO ACADEMIC PERFORMANCE

by

Marino Alexander Middleton

B.A., University of Windsor, 1968

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS (EDUCATION)

in the Faculty

of

Education

SIMON FRASER UNIVERSITY

June 1979

APPROVAL

Name: Marino A. Middleton

Degree: Master of Arts (Education)

Title of Thesis: An Exploratory Study of the Relationship of Study Habits, Expected Grade and Instructor Effectiveness Ratings to Academic Performance

Examining Committee

Chairman: M. Manley-Casimir

R. W. Marx
Senior Supervisor

P. H. Winne
Second Member

J. W. George Ivany
Dean of Education
Simon Fraser University
External Examiner

Date approved June 15, 1979.

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis or dissertation (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this thesis for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis/Dissertation:

AN EXPLORATORY STUDY OF THE RELATIONSHIP OF STUDY HABITS, EXPECTED GRADE AND INSTRUCTOR EFFECTIVENESS RATINGS TO ACADEMIC PERFORMANCE

Author: _____

(signature)

Marino MIDDLETON

(name)

June 15/79

(date)

ABSTRACT

An exploratory investigation of relationships among students' ratings of instructor effectiveness, students' study habits, and expected grade in relation to final grade is described. Two volunteer university instructors in the Faculty of Education and their undergraduate education classes participated. Students filled out questionnaires related to study habits, expected grade and instructor effectiveness approximately six weeks after the beginning of the semester. One class received five mini-tests and were asked to produce a term project during the semester. Both classes were assigned end of semester grades. Five hypothesized relationships among the above variables were presented. Separate principal component factor analyses of the instructor rating form for each class indicated that the two groups of students perceived their instructor differently. Using factor scores as predictors in a multiple regression analysis, 15% of the criterion variance in a global rating of teacher effectiveness could be determined in class A by a combination of a presentation factor and a student support factor. In course B, only 5% of the criterion variance could be determined by one factor, a skill factor. With single items as predictors, two items, "respectful of views other than his own" and "objectives of the course are clear", predicted 23% of the criterion variance for instructor A. For instructor B, 3 items, instructor style is entertaining, "instructor has good knowledge of subject matter" and "instructor speaks clearly and distinctly", determined 39% of the criterion

variance. Using a subset of items with factor loadings of .50 or greater to represent each of the student rating factors to predict teacher effectiveness, 35% of the variance on the criterion item could be accounted for by a presentation factor for instructor A. For Instructor B, 42% of the variance could be accounted for by two factors, a skill factor and style of presentation. Expected grade was found to be the best predictor of final exam mark or final grade for the class of Instructor A, accounting for 17% of the variance in grades and 27% of the variance in final exam marks. Students' end-of-semester ratings of Instructor A, who administered five quizzes throughout the term and assigned and end-of-semester paper or project, were significantly influenced by these evaluation activities. The results indicated that 54% of the variance of the end-of-semester ratings could be determined from the marks students received on the quizzes and paper/project. Discriminant analysis of the data indicated that the two groups of students differed statistically in how they rated their respective instructors and in the grade they expected in the course. It was possible to correctly classify a total of 67.6% of the students into their respective classes on the basis of instructor ratings and expected grade, a 17.6% improvement over chance classification. The findings of this study suggest that although students perceive their instructors somewhat differently, they tend to use similar aspects of teaching behaviour to judge the effectiveness of their instructors, namely, preparedness and aspects relating to presentation. The

degree to which these factors are related to global effectiveness, however, differs from one instructor to another. The findings indicate that the instructor rating scales can provide information to instructors about dimensions of performance that students use to judge instructor effectiveness. However, comparison among instructors may not be valid because of variation in the criteria students use to make the judgment from one instructor to another. In addition, this study found that grades received during the term significantly affected the ratings students give the instructor, moreso than ratings obtained near mid-term. This study examined only the relationships that may exist between instructor ratings, expected grade, study habits and grades. Future researchers may wish to explore specific hypotheses concerning these variables and experimentally determine the nature of these relationships more fully.

ACKNOWLEDGMENTS

The writer wishes to express his sincere appreciation for the generous help and guidance given by Dr. Ronald W. Marx and for the willing assistance given by Dr. Philip H. Winne.

In addition, the writer wishes to express his thanks to Dr. Michael Manley-Casimir, Director, Graduate Programs in Education, for his co-operation in carrying out this study. The co-operation and efforts of the undergraduate students who filled out the questionnaires and the graduate students who assisted in their collection is also gratefully acknowledged. Finally, the author wishes to thank Dr. W.G. Davenport for his assistance in pilot testing the instruments pertaining to this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

PURPOSE OF THE RESEARCH

A major concern in instructional research has been how to judge the effectiveness of teaching. The troubled history of process-product research and its attempts to isolate aspects of effective teaching may be found in detailed reviews (Rosenshine and Furst, 1973; Heath and Nielson, 1974 and Dunkin and Biddle, 1974). These reviews suggest that there are few consistent findings to aid teachers' decisions as to what they should or should not do while teaching in their classrooms.

Much of the early research in this area tended to borrow both variables and paradigms from other disciplines, primarily psychology. The major problem associated with this approach is that psychology has concerned itself mainly with theories and data relevant to questions of learning, and these have been considerably less applicable to the classroom than theories and data relevant to the problems of teaching. There may be two reasons for this gulf. Theories of learning typically have been developed and tested under controlled laboratory conditions involving unrealistically restrictive parameters. Second, theories of instruction have not been adequately emphasized. Brophy (1974) outlines a more detailed reason for this gap:

As a sometimes student of the history of science, I have observed that the development of a specific science usually passes through several stages. The first features heavy stress on attempts to define the scope of the science and to identify the phenomena to be subsumed under it, the kinds of research paradigms to be used, and contexts in which research is to take place. This stage was a major problem early in the science of psychology, for example, in its attempt to stake out a unique territory between philosophy on the one hand and zoology and other biological sciences on the other. This has been less of a problem with education and educational psychology, because, at least until recent innovation of new instructional methods and settings, it seemed fairly obvious that educational research should focus on teachers and students in the classrooms.

Following this first stage there usually comes a second stage of naturalistic observation and the development of taxonomies. Here, scientists attempt to identify the relevant variables in the contexts in which they work, to agree on definitions and measurement methods and to develop some baseline data which describes the typical or normal situations. This stage is then followed by a stage of correlational research in which the parameters which have been identified in the context being studied are related to one another, and information becomes systematically integrated as theories are developed. The next stage involves systematic manipulation of some of the variables to see if these experiments will produce the expected results, thus moving from correlation to causality. Finally, when applied aspects of the science are involved, the previous body of established data is drawn upon in order to diagnose and prescribe treatments for problems.

In my opinion educational research has suffered from premature closure regarding parameters and a tendency to stress trivia over more fundamental matters because the important stage of naturalistic observation and identification of the relevant and important parameters was largely neglected. (Brophy, 1974a, p. 48)

However, a gradual shift toward investigating problems relevant to instruction appears to be under way (see, for example, Walberg, 1969; Moos, 1974; Brophy and Evertson, 1974; Goldfried and D'Zurilla, 1973; Anderson, 1970; Marx, Martin, Ellis, & Hasell, 1978). However, some of the methodology still inappropriately uses unrealistic controlled laboratory conditions and the application of static experimental paradigms to classrooms where dynamic interactions predominate. Brophy (1974a, p. 48) suggests:

> This not only makes the laboratory paradigm difficult to apply; it also leads to trivial research results or to results which have little if any external validity despite their impressive internal validity (reliability, reproducibility).
>
> In short, even though it is undeniably messy and often expensive, and even though at present it usually requires working at the correlational rather than the experimental level, I advocate a return to the real, typical classroom as the setting for educational research. Also, given the present state of the field, I think we need more hypothesis-generating research (done in the naturalistic setting), withholding hypothesis-testing research until it can be more reasonably carried out. The kinds of hypothesis-testing research going on at present are largely premature, because the hypotheses being tested are rather trivial. Thus while the educational research being carried out is generally of high quality from the standpoint of experimental design and data quality, I find that I must agree with the opinions of the average teacher who tends to have little interest in the work of educational researchers.
>
> The problem is not quality; it is relevance. By and large, we simply are not studying problems that are related to the needs of the classroom teacher.

It is for reasons such as these that there have been more and more studies examining the relationships of a variety of environmental and instructional variables to academic performance in the typical classroom. Commonly, these studies have used a correlational rather than an experimental approach (e.g., see Kulik and McKeachie, 1975).

Judging the effectiveness of teaching has been an important concern in education. Faculty ratings are now used by colleges and universities to attempt to assess this parameter of teaching. Although there is widespread use of student ratings of faculty, two opposing viewpoints have been presented concerning the validity of these measures. One viewpoint is that these ratings are among the most valid of measures, since students have observed the teaching over the length of the course. Another viewpoint holds that students are not objective or mature or knowledgeable enough to judge instructional behaviour. As Brown (1976) has stated the problem:

> Of course the main question about bias arises because of grades. Can students rate teaching effectiveness validly? Or do they use their ratings as a payoff in a sort of game between teachers and students, in which teachers reward or punish students with grades and students respond in kind with ratings? If student ratings are indeed biased by grades, to what extent are they biased? (Brown, 1976, p. 573)

In a recent review, Kulik and McKeachie (1975) cite a number of studies that related student achievement as measured by final grade to student ratings of the instructor. If a

strong proportional relationship is found, it could be interpreted that students tend to rate their instructor more highly when they receive or expect a good grade. However, Kulik and McKeachie found inconsistent results in the review of studies examining this relationship.

Study habits of students may be relevant to the prediction of grades because it is possible that students' grades may be related to their study habits. That is, students with poor study habits may obtain lower grades than those students with better study habits. The importance of the relationship between grades, instructor ratings and study habits has not been determined.

For research on teaching effectiveness to become more conclusive a multivariate approach appears necessary. Appropriate information on students, teacher and the learning situation should be obtained (Peterson, 1976).

Kerlinger and Pedhazur (1973) state:

> . . .one of the reasons that teacher effectiveness studies have been relatively ineffectual for so many years is the almost purely ad hoc atheoretical nature of the thinking and research in the field (Getzels and Jackson, 1963; Mitzel, 1960; Ryans, 1960).
>
> There are effective teachers. And something makes them more effective than other teachers. In other words, there must be an explanation of teaching effectiveness. Accepting this assumption, what is probably needed in addition to a multivariate approach, is psychological and social psychological theory that can be developed, tested, and changed under the impact of empirical research and testing (Getzels and Jackson, 1963).

With a problem so complex, with so many possible variables interacting in unknown ways, the only guide to viable research will be theory - with multivariate technical methods to back up the theory. The extraction of factors from the intercorrelations of items, the calculation of factor scores, and the application of discriminant analysis may well help to solve the problem.

Statement of the Problem

The study is designed to provide data which may be helpful in suggesting answers to the following questions:

1. Which factors of student ratings contribute most to student perceptions of overall teacher effectiveness?
2. What is the relative contribution of factors of student ratings of instructors and student perceptions of overall effectiveness, in combination with expected grades, toward:
   (a) the prediction of students' achievement as measured by final exam performance;
   (b) the prediction of students' overall grades in a course.
3. What influence does knowledge of feedback on assignments have on student ratings of teaching?

CHAPTER II

REVIEW OF RELATED RESEARCH

The following review of the literature on student perception of teaching, study habits, and the relation of these to students' grades or academic success is limited to studies using what Rosenshine (1970) has termed "high inference" measures. Rosenshine calls these measures "high inference" because the rater has to make a number of inferences as to what constitutes "good" or "effective" performance. In contrast, Rosenshine used the term "low inference" to describe questionnaire or rating items that focus upon specific and relatively objective characteristics.

There are many more studies of this latter frequency-counting type (Costin, 1968; Duncan and Biddle, 1974) than there are studies of the "high inference" type. Anderson and Walberg (1974) have pointed out that "high inference" measures may be more valid than the "low inference" measures because counts of praise or questions only measure quantity, not quality. For example, a single profound and appropriate question may inspire learning more than ten superficial ones. This allows the possibility that a large number of students who rate high inference teaching behaviours may provide more valid judgments

about the quality of instruction than a researcher who counts these behaviours.

Teacher Rating Forms

College instructors are frequently evaluated by students responding to a teacher rating form. Student ratings of faculty are now in wide use. In a recent review of the literature on student ratings, Costin, Greenough and Menges (1972) summarized studies which attempted to assess criteria students used in instructor evaluations. They indicated (p. 53) that the most commonly mentioned attributes of excellent instruction as rated by students were preparedness, clarity,and stimulation of students' intellectual curiosity.

Kulik and McKeachie (1975) in their review of studies dealing with student ratings found that the factors of teaching quality identified in these studies correspond to four of the six factors found by Baerwald, Hofeller, Isaacson, Lin, McKeachie, Milholland and Zinn (1964). The four factors were Skill, Rapport, Structure and Overload. The Skill factor had the appearance of a general factor in that it correlated positively and strongly with ratings of "all-around teaching ability." They also stated that the most important influence on student ratings might be the general disposition of the student toward the instructor and toward courses in general.

Teaching conditions also appear to influence the ratings instructors receive from students. Gage (1961) reported

substantive and statistically significant differences in students' ratings of instruction between lower-level and high-level courses, elective and required courses, on-campus and off-campus courses, and smaller and larger classes. Although Kulik and McKeachie (1975) cite some evidence for differences among disciplines in ratings as indicated by Rayder (1968), they do not provide any evidence that there may be differences in attractiveness of courses given within a single department.

Many kinds of teacher characteristics have been examined for their relation to students' ratings of teaching quality. Isaacson, McKeachie and Milholland (1963) related selected personality characteristics of teachers, as assessed by self-reports from instructors and peer group nominations, to student ratings of the overall ability of teaching fellows in introductory psychology. The results indicated that instructors rated as effective had personality structures which were described as artistic, polished, effectively intelligent and imaginative. McKeachie, Lin and Mann (1971) reported the results of a number of studies which examined the relationship between teacher warmth and effective teaching as indicated by student achievement. The findings indicated that in some courses teacher warmth correlated positively with student achievement, while in other courses this relationship was negative. Downie (1952) found that knowledge of subject matter, interest in the subject, being prepared for class and motivating students to do their best were important factors proportional

to positive ratings of instructors. Gadzella (1968) asked a group of students to list criteria they would use for selecting the ideal professor. The four most important criteria selected were knowledge of subject matter, giving well planned and organized lectures, enthusiasm and interest in teaching, and a student orientation and willingness to assist outside the classroom.

Sheffield (1974) found in a survey of 24 Canadian universities that the qualities required to be a good instructor mentioned most often were being well prepared for class and presenting the material in an interesting way. Finally, Whitely and Doyle (1976) found results to suggest that students' implicit theories of teaching influence both the rating process and the dimensions of teaching which are identified from factor analyses of rating data. Their results suggest that students use a common frame of reference to evaluate items within the same category, but that the frame of reference varies between categories. In addition, the implicit theories are based on previous experience with instructors rather than error. They recommend that students' implicit theories should be maximized in the construction of student rating instruments. In this way distinctions between teaching behaviours required of the rater will more accuratedly represent his frame of reference for evaluating the teaching.

Although these studies provide information regarding the important characteristics students perceive effective

instructors to possess, they do not reveal the relative weighting or importance students assign to these factors when making decisions about effectiveness. Very few studies have examined the dimensions of teacher ratings to determine which aspects of these dimensions students use to make global judgments regarding the effectiveness of instruction or the quality of the learning environment.

Questions about the criteria and information humans use in their judgments has been a subject of investigation in research on decision making. Because students are requested to respond to a large number of items basically to form a judgment regarding the quality of instruction, rating of teachers is similar to tasks found in decision making experiments.

It is known from research in decision making (Smedslund 1955; Bruner, Goodnow and Austin, 1956; Hovland, Jenkins and Shepard 1961; Shepard 1963) that there exists an attentional phenomenon in making an evaluative judgment which allows a subject to take account of only a very limited number of factors at any one time. In addition, the research indicates that subjects often seem to lack insight into the extent to which they have restricted themselves in this way (Pollock 1962; Hoffman 1960; Balke, Hammond and Meyer, 1973). It may be that students rate instructors in a similar way, i.e. using only a few factors to make their judgments.

Instructor Ratings and Student Grades

The Costin et al. (1972) review of the literature indicates that while a large number of prior studies found significant positive correlations between grades and ratings, many others found no relationship. A variety of other more recent studies have provided further evidence. While the results are inconclusive, these studies suggest that there may be a linear relationship between grades and ratings (Holmes, 1972; Rodin & Rodin, 1972; Bausell & Magoon, 1972). Kulik and McKeachie (1975) conclude in their review of the studies in this area that the most impressive thing about studies relating students' achievement to their ratings of instructors has been the inconsistency of the results, Very high positive, very high negative, moderately positive, and near zero correlations have been reported.

Study Habits

In attempts to predict academic ineffectiveness among college students, Brown and Holtzman (1953) constructed a questionnaire which surveyed students' study habits, attitudes and motivation toward academic work. A later form of the instrument that specified several subscales that contribute to a total score was developed. Brown and Dubois (1964) found a significant correlation between Survey of Study Habits and Attitudes (SSHA) scores and grades with engineering students,

but failed to find significant correlations with humanities and science students. On the basis of data obtained from several different college populations, Brown and Holtzman (1966) found that correlations between the SSHA scores and grades increase slightly as a function of student college experience.

Goldfried and D'Zurilla (1973) tested the validity of the SSHA against criterion measures reflecting effective study behaviour. Using peer ratings and self ratings of academic effectiveness, they found that validity coefficients were higher than when grades were used as the criterion. No significant correlations were found between the SSHA and ratings of interpersonal effectiveness, indicating that the test possessed good discriminant validity.

Goldman and Hudson (1973) found that groups defined by levels of grade point average differed significantly on study strategies but not on abilities as measured by tests from the Kit of Reference tests for Cognitive Factors (French, Ekstrom & Price, (1963), possibly due to truncation of variance as a consequence of the college admission policy. They state:

> . . .the findings in the present study lend support to the idea that strategies may be more fundamental determinants of academic success than are abilities.

In addition, they continue:

> There was an absence of interaction between grade point average and major field for either abilities or strategies. This finding would suggest. . .that strategies necessary for academic success are relatively consistent across major fields. (Goldman & Hudson, 1973, p. 364).

Summary

The literature on student ratings suggest that college students equate effective teaching with three broad clusters of instructor attributes: knowledge of subject matter, organization of that subject matter for a clear and logical presentation and demonstration of an interest in the subject matter. These clusters indicate that students emphasize strong teaching skill and subject matter knowledge in their conception of effective college teaching. The review of the literature, however, identifies only attributes that students subjectively report as being important for effective teaching. There is an absence of studies that examined the specific correlates of instructor effectiveness such as style of presentation or organization, and how the factors identified actually contribute to conceptions of effective instruction.

The relationship between instructor ratings and grades does not seem to be clear. The literature indicates that in some cases there is a positive relationship whereas in other cases there is no relationship. In the studies examining this relationship, instructor rating forms were always given toward the end of the term. As a result students' ratings of instructors may be significantly influenced by knowledge of their performance during the term, for example, as communicated by quiz grades. Students who obtain poor grades during the term may attribute this to the quality of instruction and provide low ratings.

The literature on study habits indicates that there is a significant relationship between SSHA scores and grades for students in engineering programs, but there may be no relationship for other disciplines. The relationship between the SSHA and grades becomes stronger as a function of college experience. But, while good grades may often result from effective studying and good attitudes toward academic work, this is not always true. Other factors such as the nature of the course, aptitude of the student and outcome measures may also play significant roles in the determination of a relationship between study habits and grades.

The preceding review of the literature on instructor rating scales indicates that more research is required to determine the dimensions of instructor rating scales that contribute to students' judgements of instructor effectiveness. In addition, the inconsistent results relating students' achievement and their ratings of instructors requires examination. The relationship between students' study habits and students' ratings of instructors has not been explored. It may be that students with poor study habits, who consequently obtain poor marks, may rate their instructors low. Finally there may be an influence on instructor ratings because students obtain knowledge of results prior to responding to the instructor rating form. These issues are examined in this study.

CHAPTER III

METHOD

This chapter describes the procedures used in carrying out the present investigation. Specific attention is given to sampling procedure, experimental design, materials, pilot testing, administration and scoring of instruments, and analysis of data.

Sampling Procedure

For the administration of the final version of the questionnaires, two university instructors teaching introductory undergraduate education courses at Simon Fraser University volunteered their respective classes to participate in this study. As a result, 72 students from one instructor, hereafter called Instructor A., and 82 students from another instructor, hereafter called Instructor B, participated in the study.

Materials

The Instructor Rating Form for the pilot study consisted of statements adopted from Yonge and Sassenrath (1968). The form contained 21 items dealing with teaching behaviour

and 11 items focusing on aspects of the course. The statements were presented as five point rating scales relating to frequency of occurrence. Statements were arranged in random order. A pilot test of this instrument was carried out using the responses of 25 students from a community college psychology class. As a result of pilot testing 6 of these statements are not included in the final questionnaire because of low item-total test correlations (below .30). The final Instructor Rating Form consisted of 26 items and one global item for rating overall teacher effectiveness. A copy of the Instructor Rating Form including item-subtest correlations and internal consistency reliability estimates may be found in Appendix A.

The Study Habits Quesionnaire was adopted from Brown and Holtzman (1966) and consists of three areas: study organization, study techniques and study motivation. Each area of study habits contained 20 items. However, as a consequence of a pilot test of these items a number of items which had low correlations (below.30) with their respective subtests were dropped from the questionnaire. In addition a five point rating scale was substituted for the dichotomous scale used by Brown and Holtzman (1966). A copy of each of the three Study Habits Questionnaires and the internal consistency reliability estimates can be found in Appendix B.

After poor items had been removed from the tests on the basis of the small sample pilot measurements, item-test and internal consistency reliabilities were again calculated on

larger samples found in classes of Instructor A and B. For the sake of brevity of presentation only estimates calculated on Instructor B's class is presented in the appendices. Instructor A's results on these measures approximated those of instructor B. The internal consistency estimates were .95 for the Instructor Rating Form, .75 for the Study Motivation Survey, .85 for the Study Organization Survey and .73 for the Study Techniques Survey. All of these estimates indicated that the measuring instruments had acceptable internal consistency reliabilities.

Expected grade was assessed by the question: What letter grade do you ~~expect~~ in this course? The measure of performance during the semester in Class A was the total of five mini-test scores and scores on three short papers administered throughout the semester. The measure of achievement was the final exam mark for Class A at the end of the semester. The measure of grades in Class A was end-of-semester grades converted to numerical equivalents. For Class B it was semester-end grades converted to numerical equivalents. The following equivalents were used: $A^+ = 10$, $A = 9$, $A^- = 8$, $B^+ = 7$, $B = 6$, $B^- = 5$, $C^+ = 4$, $C = 3$, $C^- = 2$, $D = 1$, $F = 0$.

Design and Procedures

Data on the Study Habits and Instructor Rating Form were collected approximately six weeks after classes had started. Class A had taken and received marks on one short paper and one

quiz at the time the forms were administered. The instructor rating form was administered to Class A a second time immediately after students wrote the final exam. At this time, the students had received feedback on all five mini-tests and all three short papers, but not on the final exam.

In Class B the instructor did not personally assign grades, but left this responsibility to his teaching assistants. A second administration of the instructor rating form was not obtained in Class B, since the course had finished earlier than the author had anticipated and therefore missed the opportunity to carry out a second administration.

Data Analyses

Reliability estimates were calculated separately on each of the three Study Habits subscales and the Instructor Rating Form on Class A and Class B. A test-retest stability coefficient was calculated for each of the instructor rating items using Class A.

In order to determine what dimensions the instructor rating form was measuring, the data for each instructor was analyzed separately by a principal component factor analysis with orthogonal rotation (VARIMAX).

In order to determine the best predictors of students' global rating of instructor effectiveness, a multiple stepwise regression analysis was carried out separately for each instructor.

The criterion variable in each case was the rating given by the students on the global item ("The instructor is an effective teacher") of the Instructor Rating Form. The predictor variables in each case were the factor scores for each of the three factors generated by the separate factor analyses, a procedure advocated by Kerlinger and Pedhazur (1973). The factor scores calculated on the orthogonal factors would not be correlated with each other, thereby overcoming the multicollinearity problem which would make interpretation difficult. However, nonsalient loadings, i.e., loadings near zero, would contribute error variance to the factor scores. As a result other procedures to approximate the underlying dimensions of the Instructor Rating Form were also carried out.

As one alternative approximation procedure, the single variable that loaded most heavily on a respective factor was chosen (Gorsuch, 1974), instead of computing a complex weight matrix in which every variable was used in estimating each factor. The variable was chosen not only on the basis of high correlation with its respective factor and low correlation with other factors, but also because the qualitative nature of the variable was congruent with the apparent qualitative attribute of the factor. Using single items that have low intercorrelations with other factors avoids the problems associated with multicollinearity (Kerlinger & Pedhazur, 1973) and may even increase the predicted variance, if the items correlate high with the criterion.

Another approximation procedure created factor defined subscales by adding together students' ratings on the subset of variables that loaded at least .50 on each factor (Gorsuch, 1974). The choice of this cutoff was based on following rationale advocated by Gorsuch (1974);

> The general approach is to have a sufficiently large sample so that anything that would be of interest for interpretation would be significant. A rough check can be provided by doubling the appropriate standard error and assuring that the minimum salient loading is significant by that rough test. For example the minimum significant correlation coefficient ($p < .05$) with an n of 100 is about .2; therefore, only elements of the factor structure matrix (S) greater than an absolute value of .4 would be interpreted if the analysis was based on 100 individuals. If it was expected that elements as low as .3 would be interpreted, then a minimum n of 175 would be needed.

It can be seen that because of sample sizes of 72 and 82 in this study, it is reasonable to use .50 as the lower bound of salient loadings since twice the standard error of a correlation based on samples of these sizes are .5 and .4 , respectively. Gorsuch (1974) has indicated that the approximation procedure outlined above correlates about .8 with the more complicated procedures.

To examine the relationship between study habits, expected grades and instructor ratings as predictors of final grades, multiple stepwise regression analyses were computed separately for each instructor. As a preliminary data reduction procedure, the Instructor Rating Form responses were combined into a total

score by simply summing item scores. This seemed appropriate because total scores of the salient loaded items on the three factors of the Instructor Rating Form were highly intercorrelated. This would make interpretation of the regression analyses difficult. The three subtests for study habits were also summed to yield a total score. Because the statements were worded in a manner that would indicate poor study habits if answered in the affirmative, high scores indicate poorer study habits.

To examine the relationship between expected grade, study habits, and instructor ratings as predictors of achievement measured by final mark in the end-of-semester exam, separate multiple stepwise regression analyses were computed based on the data for Class A.

The first analysis used the total score on the first administration of the instructor rating form as the dependent variable and the total score of the study habits' subtests, expected grade and a total score of a combination of mini-tests and a term paper as independent variables. The second analysis used the total score on the second administration of the Instructor Rating Form (immediately upon completion of their final exam) as the criterion variable and the total score of the study habits subtests combined, expected grade and total score of mini-tests combined with the term paper mark as predictor variables.

As a final analysis to determine if the two Classes A and B differed significantly on the basis of total scores on

study habits, instructor ratings or expected grades, a stepwise discriminant analysis was carried out using these variables to predict class membership.

CHAPTER IV

RESULT AND DISCUSSION

Table 1 presents a summary of statistical information and an intercorrelation matrix of the variables for Instructor A. Table 2 presents similar information for Instructor B.

Results of the Factor Analyses

In order to determine the number of factors for rotation, scree tests (Cattel, 1966) were carried out on the data from each class. A basic rationale for the scree test is that the battery of variables on the Instructor Rating Form measures a limited number of factors well and a larger number of trivial, specific and error factors much less well. Therefore the predominant factors account for most of the variance and are large, whereas the other remaining factors are more numerous but each accounts for only a small portion of the total common variance. Since the principal component solution extracts factors by size, the factors accounting for the larger amounts of variance are extracted first and then trivial factors are extracted later. Because the "smaller variance" factors are usually numerous and are extracted in order of size, plotting the amount of

Table 1

Summary Table of Means, Standard Deviations, Intercorrelations And Reliabilities For Instructor A

	Variable	Mean	S.D.	Correlations											
				1	2	3	4	5	6	7	8	9	10	11	12
1.	Item 27 1st	4.09	1.07		29*	-09	17	06	00	68**	38**	60**	54**	50**	-12
					47	*52*	*50*	*64*	*70*	*70*	*70*	*70*	*70*	*70*	*70*
2.	Item 27 2nd	2.99	2.19			02	02	12	05	33**	64**	20	27*	29*	59**
						49	*48*	*44*	*49*	*49*	*49*	*49*	*49*	*49*	*49*
3.	Final Exam	37.52	5.33				65**	52**	-15	-08	03	02	-02	-19	22
							52	*49*	*54*	*54*	*54*	*54*	*54*	*54*	*54*
4.	Final Grade	6.42	2.54					41*	-14	03	14	06	10	-10	21
								47	*52*	*52*	*52*	*52*	*52*	*52*	*52*
5.	Expected Grade	5.65	1.36						-08	08	23*	06	08	09	10
									66	*66*	*66*	*66*	*66*	*66*	*66*
6.	Study Habits	111.39	18.01						(88)	02	-08	00	-01	09	-01
										72	*72*	*72*	*72*	*72*	*72*
7.	Instructor Rating 1st	109.26	13.39							(88)	44*	92**	87**	86**	-14
											72	*72*	*72*	*72*	*72*
8.	Instructor Rating 2nd	82.10	49.14								(90)	41**	45**	35**	-28
												72	*72*	*72*	*72*
9.	Presenta-tion Factor	36.47	7.22									(86)	71**	69**	-11
													72	*72*	*72*
10.	Prepara-tion Factor	41.50	9.09										(75)	71**	-09
														72	*72*
11.	Support Factor	19.72	5.82											(77)	-02
															72
12.	Achievement	56.37	43.91												

Note: Alpha reliability coefficients are in parentheses; number of cases are directly under correlation coefficients; decimals are omitted on correlations.

* $p < .05$
** $p < .01$

Table 2

Summary Table of Means, Standard Deviations, Intercorrelations And Reliabilities for Instructor B

	Variable	Mean	S.D.	Correlations							
				1	2	3	4	5	6	7	8
1.	Item 27	3.26	1.23		-01 *62*	-01 *69*	-25* *76*	69** *76*	66** *76*	49** *76*	51** *76*
2.	Final Grade	7.25	2.28			19 *62*	13 *68*	07 *68*	-15 *68*	14 *68*	01 *68*
3.	Expected Grade	6.12	1.48				-11 *73*	-02 *73*	-08 *73*	19 *73*	-15 *73*
4.	Study Habits	116.23	20.02				(88)	-12 *82*	-13 *82*	-07 *82*	-14 *82*
5.	Instructor Rating	95.84	18.12					(95)	94** *82*	91** *82*	90** *82*
6.	Presentation Factor	38.40	6.71						(86)	77** *82*	78** *82*
7.	Preparation Factor	34.04	4.07							(75)	78** *82*
8.	Knowledge Factor	29.46	3.95								(77)

Note: Alpha reliability coefficients are in parentheses; number of cases are directly under correlation coefficients; decimals are omitted on correlations.

* $p < .05$
** $p < .01$

common variance accounted for by successive factors on graph paper should result in a straight line with a downward slope. The larger factors should not fall on this line because they will account for a great deal more variance and because there are usually fewer of them. Overall, then, a graph of common variance accounted for by successive factors should appear like an arm with a bent elbow. Application of the test is simple. All the eigenvalues of the factors are plotted with their values along the ordinate. The factor numbers are plotted along the abcissa. The point where the factors accounting for large amounts of common variance curve above the straight line formed by the smaller roots gives the number of factors. Using the bent elbow analogy, these factors are representative of the upper arm (see Figure 1). Using Catell and Jaspers' (1967) suggestion that the number of factors be taken immediately before the straight line begins, the procedure yielded a 3 factor solution for each case.

The Kaiser-Meyer-Olkin measures of sampling adequacy were 0.97 and 0.98, respectively, for Instructor A and Instructor B, indicating that there were adequate number of cases in each of the samples to proceed with the analysis (see Dzinban & Shirkey, 1974).

Chi-square tests were used to determine if there was significant variance left after a certain number of factors have been extracted. If there is statistically significant variance left, then at least one more factor could be extracted

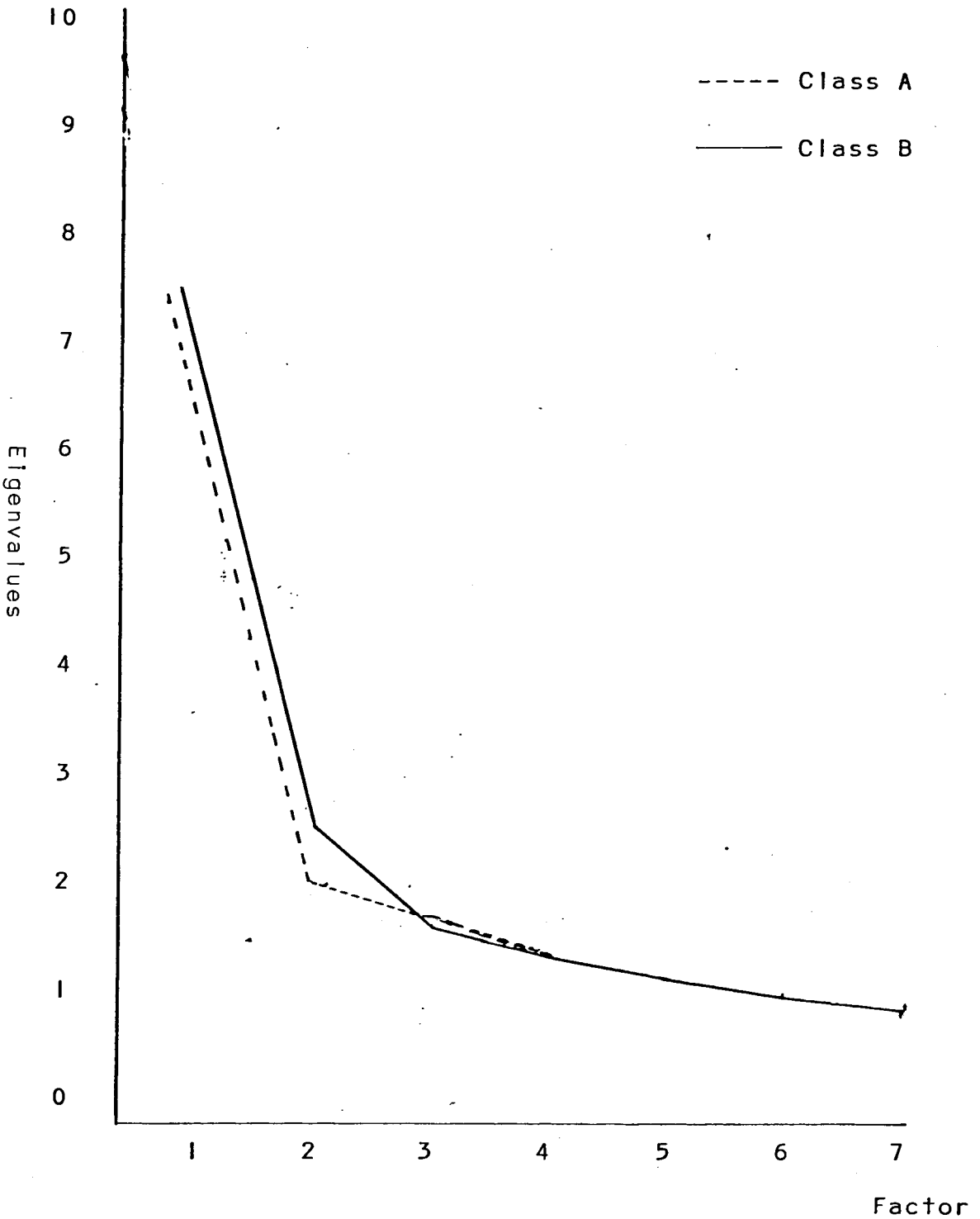

Figure 1. Scree tests for Instructors A and B.

with some assurance that non-chance variance is being processed. If there is no significant variance, then the "correct" number of factors has been extracted. The test is done by sequentially testing the total variance unaccounted for by the factors already extracted until this quantity is found to be non-significant.

The chi-square test for adequate factoring yielded values of 158.75 and 56.62 at 276 degrees of freedom when 3 factors were extracted for Instructors A and B, respectively. In each case the chi-square value was below statistical significance ($p > .05$). Thus, three factors adequately estimates the common variance factor structure in each case.

Table 3 presents the factor loading matrix for Instructor A. Table 4 presents the factor loading matrix for Instructor B. In both cases, variables are re-ordered from highest to lowest factor loadings within each respective factor.

Interpretation of the Factors

The variables loading on the first factor for Instructor A tend to describe style of presentation. The second factor deals with preparedness of the presentation, and the third factor deals with behaviours that are supportive of students. The three factors are similar to three of the factors identified by Isaacson, et al. (1964).

For Instructor B, the first factor also appears to be related to style of presentation but with an emphasis on

Table 3

Instructor Rating Form Factor Loading Matrix for Instructor A

	Item on Instructor Rating Form	Factor 1	2	3
8	objectives clear	-.70	.25	-.06
10	exhibits self reliance	-.69	.05	.32
12	exhibits all around teaching ability	-.68	.32	.37
4	speaks fluently	-.68	.32	.37
5	materials suited to needs	-.64	.42	.05
25	materials suitable to class	-.63	.51	.17
3	speaks clearly	-.56	.33	.18
16	good knowledge of subject matter	-.56	.08	.34
15	request for papers is adequate	-.49	-.01	.17
22	classes are well prepared	-.06	.65	.05
17	stimulates intellectual curiosity	-.30	.65	.37
1	appropriate techniques	-.06	.65	.09
14	presents the material in an interesting way	-.32	.64	.34
21	course contributes to general education	-.40	.59	-.02
24	instructor interested in subject	-.34	.54	.35
18	knowledgeable of new developments	-.48	.53	.17
23	Explanations are clearly presented	.12	.53	.20
20	course is well organized	-.40	.48	-.02
9	relates to other fields	-.11	.38	-.01
19	frequency of exams appropriate	-.20	.23	.13
13	respectful of views	-.31	-.15	.76
11	students feel free to express opinions	-.22	-.09	.69
7	permissive	.08	.29	.67
6	willingness to help	-.47	-.03	.63
2	humour	-.08	-.51	.61
26	style is entertaining	-.21	.50	.54
	Percent of Total Variance	35.0	7.0	7.3

Table 4

Instructor Rating Form Factor Loading Matrix

for Instructor B

	Item of Instructor Rating Form	Factor 1	2	3
26	style is entertaining	.84	.21	.01
14	presents the material in an interesting way	.75	.33	-.19
12	exhibits all around teaching ability	.71	.38	-.26
7	permissive	.71	.07	-.19
17	intellectual curiosity	.66	.41	-.17
2	humour	.64	.06	-.32
11	students feel free to express opinions	.60	.14	-.28
6	willingness to help	.54	.07	-.29
21	course contributes to general education	.54	.44	-.21
9	relates to other fields	.51	.12	.11
3	speaks clearly	.15	.74	-.07
8	objectives clear	.18	.69	-.07
4	speaks fluently	.06	.69	-.17
10	exhibits self reliance	.08	.62	-.43
22	classes are well prepared	.20	.61	-.44
1	appropriate techniques	.29	.58	-.20
20	course is well organized	.36	.52	-.26
23	explanations are clearly presented	.39	.50	-.42
18	knowledgeable of new developments	.05	.44	-.41
19	frequency of exams appropriate	-.10	.35	-.53
5	materials suited to needs	.17	.33	-.53
13	respectful of views	.36	.05	-.57
15	request for papers adequate	.20	.12	-.59
24	instructor interested in subject	.39	.12	-.61
25	materials suitable to class	.22	.19	-.73
16	good knowledge of subject matter	.36	.31	-.77
	Percent of Total Variance	37.5	9.6	6.2

stimulation or arousing interest. This first factor can be described as a skill factor because of the apparent similarity of the items to dimensions identified by Kulik and McKeachie (1975). The second factor appears to be related to a well prepared presentation, and the third factor deals with the balance between the instructor's knowledge and interest in his subject and the level of appropriateness of his presentation to the class.

Regression Analyses

Predicting Effectiveness Ratings

Table 5 presents the summary table of the separate linear stepwise regression analysis for Instructors A and B using factor scores as independent variables and item 27 (the global effectiveness rating) on the Instructor Rating Form as the dependent variable. Only variables whose F value was significant ($p < .05$) were allowed to enter into all regression equations.

The results presented in Table 5 indicate that for Instructor A the preparedness factor and the presentation factor contribute significantly toward the prediction of the global teacher effectiveness item rating. For Instructor B only the preparedness factor contributes significantly toward the prediction of global teacher effectiveness item rating. The results indicate qualitative differences regarding reliable

Table 5

Summary of Multiple Regression Analysis Using Factor Scores to Predict Item 27 on the Instructor Rating Form for Instructors A (n = 72) and B (n = 82)

Variable Entered	Multiple R	R^2	Increase in R^2	b	F
INSTRUCTOR A					
Factor 2 (preparedness)	0.34	0.116	0.116	0.37	9.20*
Factor 1 (presentation)	0.39	0.153	0.037	0.21	2.98*
Constant				4.09	
INSTRUCTOR B					
Factor 2 (preparedness)	0.22	0.05	0.05	-0.23	4.18*
Constant				4.211	

*$p < .05$

predictors of global effectiveness in teaching and quantitative differences across the two instructors in how predictive student ratings are to overall judgments about teaching effectiveness.

The use of factor scores to predict the global effectiveness item has both advantages and disadvantages. The advantage is that due to the orthogonal rotation of the principal component solution, the factor scores which are generated are uncorrelated with each other. Thus, interpretation of the multiple regression analysis is made easier due to the complete

absence of multicollinearity. The disadvantage is that additional variables, i.e., items on the rating form, that do not have salient factor loadings are contributing variance to the factor scores. As a consequence of this the internal consistency reliability estimates of the factor is lowered and in turn the predictive power diminished.

The results of the regression analysis for Instructors A and B involving the use of single items as an approximation procedure to represent each factor as the predictor are presented in Table 6.

Table 6

Summary of Multiple Regression Analysis Predicting Instructor Effectiveness Using One Variable to Approximate a Factor for Instructors A (n = 72) and B (n = 82)

Variable Entered	Multiple R	R^2	Increase in R^2	b	F
INSTRUCTOR A					
item 13	.39	.15	.15	0.34	8.25*
item 8	.49	.24	.09	0.27	7.46*
constant				1.84	
INSTRUCTOR B					
item 26	.54	.29	.29	0.41	23.78*
item 16	.60	.36	.07	0.31	5.51*
item 3	.62	.39	.03	0.15	3.02*
constant				0.43	

*$p < .05$

As can be seen from these results for Instructor A, item 13 on the Instructor Rating Form ("the instructor is respectful of views other than his own?) shares 15% of the variance of the global effectiveness item on the form. An Additional 9% of the variance on item 27 can be determined by item 8 ("the objectives of the course are clear"). This procedure resulted in a total of 24% of the variance of the teacher effectiveness item than can be determined on the basis of two items related to student support and preparation of presentation. As expected, this is an increase over the percentage of variance accounted for in Table 5 where factor scores were used.

The majority of predicted variance on the teacher effectiveness criterion for instructor B, about 29%, can be determined on the basis of item number 26 on the Instructor Rating Form (the instructor's style is entertaining") and an additional 7.4% of the variance can be determined by item 16 ("the instructor has a good knowledge of his subject matter"). Finally, a further 2.6% of the variance can be determined by item 3("the instructor speaks clearly and distinctly"). The total variance that is shared between the teacher effectiveness criterion and these 3 items is 38.7%. Again, the predicted increase in variance accounted for was obtained.

When single items that correlated highest with a factor were used to predict instructor effectiveness, approximately 24% to 39% of the criterion variance for each instructor could be determined on the basis of 2 and 3 items respectively. If the entire instructor rating form is viewed to be a measure

of instructor effectiveness, this form has construct validity. However, these results are significant in that they indicate the form is measuring different aspects of this construct from one instructor to another.

Students of Instructor A see him as a relatively effective teacher because he is respectful of views other than his own and the objectives of his course are clear. Students of Instructor B perceive him to be a relatively effective instructor because his style is entertaining, he has a good knowledge of his subject matter and he speaks clearly and distinctly.

The results of entering the total scores of a subset of variables that correlated .50 or higher with a factor to estimate a factor into regression analyses to predict the global effectiveness item 27 may be found in Table 7 for Instructors A and B.

Table 7

Summary of Multiple Regression Analysis Using a Subset of Variables to Approximate a Factor for Instructors A and B (n = 72) and B (n = 82), Predicting Instructor Effectiveness

Variable Entered	Multiple R	R^2	Increase in R^2	b	F
INSTRUCTOR A					
Presentation	.59	.35	.35	.07	40.49*
Constant				.94	
INSTRUCTOR B					
Skill	.61	.37	.37	0.08	9.04*
Preparedness	.65	.42	.05	0.37	5.22*
Constant				.01	

*$p < .05$

The results indicate that for Instructor A only one factor, presentation, predicted the global effectiveness item. Thirty-five percent of the variance of the global effectiveness item could be determined by the variance on this factor. For Instructor B, two factors, skill and preparedness, were significant predictors of the global instructor effectiveness item. The first factor determined 37% of the variance on the global effectiveness criterion item, and the second added an additional 4%.

However, some caution is in order here because Tables 1 and 2 show that the approximation factors composed of total scores on a subset of variables are highly intercorrelated for each instructor.

The interpretation of the multiple regression analysis is made difficult because differences in their zero order correlations with the criterion could result from sampling error, and the remaining predictors would not substantially increase the variance that could be determined on the criterion item.

The results of the approximation procedures and the regression analysis with factor scores as predictors tend to corroborate the results cited in the review of the literature, e.g., Kulik and McKeachie (1975), Ryans (1960) and Sheffield (1974). Students apparently make their judgments about instructor effectiveness on the basis of only a few dimensions. The

effective instructor is viewed as being well prepared for class, and presents knowledge in a way that is well organized and that increases students' interest in and appreciation of the subject.

The particular items that correlated highest with a factor reveal the differences in the student perceptions of different instructor's effectiveness. Students apparently "home in" on those specific instructor attributes that can be equated with instructor effectiveness. These specific attributes will probably vary from instructor to instructor, class to class, and course to course. As may be seen from the results of this study, entertaining style, good knowledge of subject matter and clarity of speech tended to be the criteria for one group of students and one instructor. For another group of students and a different instructor, respect for other people's views and clear objectives tended to be criteria. But since these two courses serve similar roles in the Faculty of Education undergraduate curriculum and many students eventually take both courses, it is reasonable to conclude that much of the differences in these results reflect differences in instructors, perhaps moreso than differences in students.

An important implication of the results is that when different instructors are rated with the same form by different students, the criteria used in the judgment process to assess effective instruction vary. If instructors wish to improve their instruction based on student ratings, they should determine which dimensions of rating forms their students use to make

judgements of instructor effectiveness. If a number of classes from different disciplines are analyzed separately, there may be common factors that contribute to the judgment of effective instruction. Further research along the design of this study appears warranted to determine these dimensions.

After identifying the dimensions that contribute to the rating of effective instruction, the instructor who wishes to improve his ratings can modify his instructional techniques accordingly. If techniques outlined in this study are combined with those mentioned by Aleamont (1978) it may be possible to provide valid, reliable and meaningful feedback to instructors. As Aleamont has mentioned, providing computerized results of instructor evaluations along with consulting sessions with another faculty member concerning that information would allow opportunity to discuss and explore possible ways of rectifying areas of instructional weakness as perceived by students. Perhaps the previous shortcomings of student ratings of instructors for improving instruction have resulted because feedback is either too delayed, too global or the judgmental process is ill understood (e.g., Kulik and McKeachie, 1975).

Predicting Grades

The results of the regression analysis to examine the relationship between expected grade, study habits and instructor ratings and achievement as measured by final mark in the end of semester exam for Instructor A are reported in Table 8.

Table 8

Summary of Multiple Regression Analysis to Predict Final Exam Mark Using Expected Grade, Study Habits and Instructor Ratings for Instructor A

Variable Entered	Multiple R	R^2	Increase in R^2	b	F
Expected Grade	.52	.27	.27	2.05	17.68*
Constant				25.97	

*n = 72 p < .05

These results indicate that only one of these variables significantly determines the variance of end-of-semester exam mark. This variable is expected grade and it determines 27% of the variance of the exam mark.

The results of the relationship between study habits, expected grade, instructor ratings and final overall grade at the end of the semester for Instructor A may be found in Table 9.

Table 9

Summary of Multiple Regression Analysis to Predict Overall Grade Using Expected Grade, Study Habits and Instructor Ratings for Instructor A

Variable Entered	Multiple R	R^2	Increase in R^2	b	F
Expected Grade	.41	.17	.17	.76	9.12*
Constant				2.11	

*n = 72 p < .05

These results indicate that expected grade is the only significant predictor of final grade and that for Instructor A, about 17% of the variance of final grades can be determined by expected grade. The results for Instructor B for this analysis indicated that none of the independent variables were significant predictors of end-of-semester grade.

The best predictor of final overall grade or final exam mark in this study for instructor A was expected grade. This is consistent with some of the literature (e.g. Walsh, 1967), who found a close correspondence between the grade which a student said he expected and the actual grade received. These results may indicate that those students who expect higher grades may be more motivated to attain the higher grade. This hypothesis requires verification.

There were no significant results in these analyses for Instructor B. The explanation for this remains unclear, but it may be possible that more students in his class were unable to estimate their final grade than was the case in Class A. It is also possible that the grading procedure was different from those which the students were familiar with.

The finding that there was no significant relationship between reported study habits and grades in this study is congruent with the Brown and Dubois (1964) results which failed to find significant correlations among these variables with humanities and science students.

The regression analysis results for Instructor A to ascertain what percentage of the variance on the second administration of the Instructor Rating Form (at the end of semester) can be determined by expected grade, study habits or marks received on mini tests and major paper (achievement) may be found in Table 10.

Table 10

Summary of Multiple Regression Analysis for Instructor A to Predict Responses on Second Administration of Instructor Rating Form Using Expected Grade, Study Habits and Achievement on Quizzes and Project

Variable Entered	Multiple R	R^2	Increase in R^2	b	F
Achievement	.74	.54	.54	0.82	76.04*
Constant				35.6	

*n = 72 $p < .05$

These results indicate that 54% of the variance of the global score on the Instructor Rating Form at the end of the semester can be determined from the marks of a composite score of mini tests and major paper (achievement).

An interesting question allied to student ratings of instructors and grades is whether students tend to rate courses more highly when they either obtain or expect good grades in them. Do students use their ratings as a payoff in a sort of game between teachers and students in which students perceive

that teachers reward or punish students with grades and students respond in kind with ratings? If this is actually what occurs, to what extent are the ratings influenced by these evaluations? Previous literature on this issue has been divided. Some researchers have found no relationship, others have found significant but low correlations, around .30, between grades and ratings (e.g., Costin, Greenough and Menges, 1968).

The results of this study indicate that student grades on quizzes and term paper or project are significantly related to instructor ratings. The high positive correlation indicates a strong positive, linear relationship between marks obtained during the course and subsequent instructor rating by students at the end of the course. The data analyses indicate that Instructor A's average ratings dropped from 109.26 six weeks into the course to 82.09 at the end of the course. The strong positive relationship between grades and the Instructor Rating Form scores indicates that as grades rise there is also a rise in the Instructor Rating Form score.

Further research of a controlled experimental nature would be required to ascertain whether or not causal inferences are warranted in the relationship between knowledge of results or achievement measures during the term and subsequent ratings of instructors by students.

The results for Instructor B to determine the relationship between expected grade, study habits and Instructor Rating Form responses may be found in Table 11.

Table 11

Summary of Multiple Regression Analysis for Instructor B to Predict Responses on Instructor Rating Form Using Expected Grade and Study Habits

Variable Entered	Multiple R	R^2	Increase in R^2	b	F
Study Habits	-.27	.07	.07	-0.24	5.58*
Constant				124.24	

*n = 82 $p < .05$

The zero order correlation of -.27 between study habits and instructor rating indicates that as study habits get worse (higher study habit score) the instructor ratings get lower. This result indicates that people with poorer study habits tend to rate their instructors less favourably.

One possible reason that previous research has indicated either no relationship or low correlation between achievement and ratings may be due to the collapsing of scores across instructors and courses. Steps should probably be taken to determine the degree of influence student assignments such as tests, quizzes and term papers or projects have on individual instructors teaching specific courses. At least the fact that these evaluations might well be influencing ratings should be taken into account by the administrators looking at such faculty ratings, especially since some faculty members' ratings could be more severely affected by these influences than others. Unless such

considerations are taken into account ratings may tend to be counterproductive in that they would unfairly reward not the better educators but merely those who are more lenient.

Discriminant Analysis

The results of the discriminant analysis may be found in Table 12.

Table 12

Discriminant Analysis Classification Matrix and Standardized and Unstandardized Discriminant Function Coefficients

		Predicted Membership				
Actual Membership	(Missing)	Instructor A		Instructor B		Total
Instructor A	6	41	62%	25	37.8%	66
Instructor B	9	20	27.4%	53	72.6%	73

Overall classification accuracy 67.3%

Standardized Function Coefficients

-0.44 Expected Grade
0.88 Instructor Rating

Unstandardized Function Coefficients

-0.30 Expected Grade
0.06 Instructor Rating
-4.12 Constant

The discriminant analysis accurately predicted 41 out of 66 students to belong to Class A and 53 out of 73 students to Class B. The two groups were significantly differentiated on the basis of two variables expected grade, $F = 9.60$, $df = 2$ and 136 ($p<.05$) and instructor rating scores $F = 15.12$, $df = 1$, and 173 ($p<.05$). The relative size of the standardized function coefficients shows that the instructor rating is doing most of the discrimination.

It is evident from an examination of the means of Tables 1 and 2 that Instructor A obtained a higher score on instructor ratings than did Instructor B. In addition, the students of Instructor A expected a lower grade (B-) than the students of Instructor B (B). These results may account for some of the differences found by the multiple regression analyses of data on Instructors A and B (e.g. expected grades were not significant predictors of actual grade for instructor B).

Limitations of the Study

Some of the shortcomings of this study are the following.

1. There were only two instructors and therefore generalizability is limited.

2. Sampling was of only one department, a faculty of Education.

3. Only undergraduate courses were sampled.

4. There may have been contamination by the use of only teachers who volunteered to be rated.

5. There may have been contamination by some lack of anonymity for students despite being told by the author that this information would be confidential.

6. Student ability measures were not used to aid in prediction.

CHAPTER 5

CONCLUSION

This study was designed to identify aspects of teacher behaviour that students use to make judgments about effective teaching. In addition, the relationships between students' self reported study habits, expected grade, instructor ratings and end of semester grades and achievement were explored. Finally, the relationship between student achievement as measured by quizzes and term project and instructor rating was explored.

The results of the study indicate that there are both qualitative and quantitative differences in the way students in different courses perceive an individual instructor's teaching effectiveness. Although the main criteria used by students in assessing instructor effectiveness were related to aspects of preparation and presentation, other factors such as their knowledge of the subject and support of students were also used differently in judging instructors. The single item to approximate a factor when input into a regression analyses to predict the global instructor effectiveness item resulted in two items that were significant predictors for instructor A. These items were item 13 (the instructor is respectful of views other than his own) and item 8 (the objectives of the course are clear). For

Instructor B three items were significant predictors of the teacher effectiveness item. These were item 26 (the instructor's style is entertaining), item 16 (the instructor has a good knowledge of his subject matter), and item 3 (the instructor speaks clearly and distinctly).

The results indicated that in the case of instructor A's class, expected grade is significantly related to end-of-semester grade. In addition, there was a significant relationship between expected grade and final exam mark for Instructor A's class. No significant relationship was found between expected grade and end-of-semester grade for Instructor B's class. This may have been as a result of significant differences between the two classes on the basis of expected marks. Further investigation as to causes for these different expectations could be the topic of future research in this area.

The results of the study indicate that for Instructor A there was a significant relationship between the total scores made up of quizzes and term paper and subsequent ratings of an instructor. This relationship offers the possibility that students' ratings of instructors are influenced by knowledge of results of tests.

The finding that there was no significant relationship between study habits and final grades for both classes is congruent with some of the literature (e.g. Brown and Dubois, 1964), which found no significant relationship between these variables with science and humanities students.

The finding of a significant relationship between study habits and instructor rating score for Instructor B is interesting. The zero order correlation of -.27 between study habits and instructor rating scales indicates that as study habits get worse (higher study habit score), the instructor ratings gets lower. This result indicates that people with poorer study habits tend to rate their instructor less favourably but since this did not occur with Instructor A's class, generalizability is questionable.

The purpose of this study was to examine the relationship of a variety of student and instructional variables and academic performance in a university setting using a correlational rather than an experimental approach. Further research involving other classes, instructors and disciplines need to be conducted to generalize from these data and develop theories. After this the systematic manipulation of variables to establish causality would be required. Finally, the information can be applied to improve aspects of the educational process.

Bibliography

Aleamont, L.M. The usefulness of student evaluations in improving college teaching. Instructional Science, 1978, 7, 95-105.

Anderson, G.J. Effects of classroom social climinate on individual learning. American Educational Research Journal, 1970, 7, 135-152.

Anderson, G.J. & Walberg,H.J. Assessing classroom learning environments. In K. Majoribanks (Ed.), Environments for Learning, Sussex: King, Thorne and Stace Ltd., 1974.

Balke, W.M., Hammond, K.R. & Meyer G.D. An alternate approach to labour-management relations. Administrative Sciences Quarterly, 1973, 18, 311-327.

Bartlett, M.S. Tests of significance in factor analysis. British Journal of Psychology, 1950 3(2), 77.

Bausell, R.B. & Magoon, J. Expected grade in a course, grade point average, and student ratings of the course and the instructor. Educational and Psychological Measurement, 1972, 32, 1,013-1,023.

Brophy, J.E. Some good five cent cigars. Educational Psychologist, 1974, 11, 46-51.

Brophy, J.E. & Evertson, C. The Texas teacher effectiveness project. Trends in the Study of the Classroom, T. Good (Ed.) APS Press, New York, 1974.

Brown, D.L. Faculty ratings and student grades: a university wide multiple regression analysis. Journal of Educational Psychology, 1976, 68, No. 5, 573-578.

Brown, F. & Dubois, T. Correlates of academic success for high quality freshman men. Personnel and Guidance Journal, 1964, 42, 603-607.

Brown, W.F. & Holtzman, W.H. Survey of study habits and attitudes, form C, New York: The Psychological Corporation, 1966.

Bruner, J.S., Goodnow, J.J & Austin, G.A. A Study of Thinking, New York: Wiley, 1956.

Cattel, R.B. The scree test for the number of factors. Multivariate Behavioral Research, 1966, 1(2), 245.

Cattel, R.B. & Jaspers, C. A general plasmode No. 30 (10-5-2) for factor analytic exercises and research. Multivariate Behavioral Research, 1967, 3, 67.

Costin, F. Survey of opinions about lectures. University of Illinois, Department of Psychology, 1968 (Mimeographed).

Costin, F., Greenough, W.T. & Menges, R.J. Student ratings of college teaching: reliability, validity and usefulness. Review of Educational Research, 1968, 28, 1079-1085.

Downie, N.W. Student evaluation of faculty. Journal of Higher Education, 1952, 495-496, 503.

Dunkin, M. & Biddle, B. The Study of Teaching, New York: Holt, Rinehart and Winston, 1974.

Dzinban, C.D. & E.C. Shirkey. When is a correlation matrix suitable for factor analysis? Psychological Bulletin, 1974, 81, 358-561.

French, J. Social class and the relationship of problem solving styles to the factor composition of tests. Educational and Psychological Measurement, 1965, Vol. 9-28.

French J., Ekstrom, R. & Price, L. Kit of reference tests for cognitive factors. Princeton, N.J.: Educational Testing Service, 1963.

Frederiksen, C. Abilities, transfer and information retrieval in verbal learning. Multivariate Behavioral Research Monographs, 1969, 21-82.

Gadzella, B.M. College student views and ratings of an ideal professor. College and University, 1968, 44, 89-96.

Gage, N.L. The appraisal of college teaching: an analysis of ends and means. Journal of Higher Education, 1961, 32, 17-22.

Garcia, D. & Whigham, N. Validity of SSHA administered before and after college experience. Educational and Psychological Research, 1958, 18, 845-851.

Getzels, J.W. & Jackson, P.W. The teacher's personality and characteristics. N.L. Gage (Ed.), Handbook of Research on Teaching, Skokie, Ill.: Rand McNally, 1963.

Goldfriend, M.R & D'Zurilla, T.J. Prediction of academic competence by means of the survey of study habits and attitudes. Journal of Educational Psychology, 1973, 64, No. 1, 116-122.

Goldman, R.D. & Hudson, D.J. A multivariate analysis of academic abilities and strategies for successful and unsuccessful college students in different major fields. Journal of Educational Psychology, 1973, 65, 364-370.

Gorsuch, R.L. Factor Analysis. Toronto, W.B. Saunders, 1974.

Heath, R.W. & Nielson, M.A. The research basis for performance based teacher education. Review of Educational Research, 1974, 44, 463-484.

Hoffman, P.J. The paramorphic representation of clinical judgement. Psychological Bulletin, 1960, 57, 116-131.

Holmes, D.S. Effects of grades and disconfirmed grade expectations on students' evaluations of their instructor. Journal of Educational Psychology, 1972, 63, 130-133.

Isaacson, R.L., McKeachie, W.J. & Milholland, J.E. Correlation of teacher personality variables and student ratings. Journal of Educational Psychology, 1963, 54, 110-117.

Isaacson, R.L., McKeachie, W.J., Milholland, J.E., Lin, Y.G., Hofeller, M., Baerwaldt. J.W., & Zinn, K. 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.

Kulik, J.A. & McKeachie, W.J. The evaluation of teachers in higher education. Review of Research in Education. 1975, 3, 210-239.

Marx, R.W., Martin, J., Ellis, J.F. & Hasell, J. Improving the instructional effectiveness of university teaching assistants: report 1. Canadian Journal of Education, 1978, 3, 1-12.

McKeachie, W.J., Lin, Y.G. & Mann, W. Student ratings of teacher effectiveness: validity studies. American Educational Research Journal, 1971, 8, 435-445.

Mitzel, H. Teacher effectiveness. C. Harris (Ed.) Encyclopedia of Educational Research. 34 ed., New York: MacMillan, 1960.

Moos, R.H. Evaluating Treatment Environments: A Social Ecological Approach. New York: John Wiley and Sons, 1974.

Nunnally, J.C. Introduction to Psychological Measurement. Toronto McGraw Hill, 1970.

Peterson, P.L. Interactive effects of student anxiety, achievement orientation and teacher behaviour on student achievement and attitude. Paper presented at the meeting of the American Educational Research Association, San Francisco, California, April, 1976.

Pollack, I. Action selection and the Yntema-Torgerson "worth" function. Paper read at the meeting of the Eastern Psychological Association, April 27, 1962.

Rayder, N.F. College student ratings of instructors. Journal of Experimental Education, 1968, 37, 76-81.

Ryans, D. Prediction of teacher effectiveness. C. Harris (Ed.), Encyclopedia of Educational Research, 3rd ed., New York: MacMillan, 1960.

Rodin, M. and Rodin, B. Student evaluations of teachers. Science, 1972, 177, 1, 164-1,166.

Rosenshine, B. Evaluation of instruction. Review of Educational Research, 1970, 40, 279-300.

Sheffield, E.F. Teaching in the Universities, No One Way. Toronto McGill and Queen's University Press, 1974.

Shepard, R.N. Analysis of proximities as a technique for the study of information processing in man. Human Factors, 1963, 5, 33-48.

Shepard, R.N., Novland, C.I. & Jenkins, H.M. Learning and memorization of classifications. Psychological Monographs, 1961, 75 (13) (Whole No. 517).

Smedslund, J. Multiple-Probability Learning. Oslo: Akademisk Forlag, 1955.

Walberg, H.J. Social environment as a mediator of classroom learning. Journal of Educational Psychology, 1969, 60, 443-448.

Walsh, W.B. Validity of self-report. Journal of Counseling Psychology, 1967, 14, 18-23.

Whitely, S.E. and Doyle, K.O. Implicit theories in student ratings. American Educational Research Journal, 1976, 13, 241-253.

Yonge, G.D. & Sassenrath, J.M. Student personality correlates of teacher ratings. Journal of Educational Psychology, 1968, 59, 44-52.

APPENDIX A

Instructor Rating Form

INSTRUCTOR RATING FORM

n = 82

INSTRUCTIONS

Please circle the appropriate number beside each statement. Below is an explanation of the ratings:

1 = Never
2 = Not often
3 = Fairly often
4 = Frequent
5 = Very frequent

		item-subtest correlation					
1.	The instructor uses appropriate techniques of instruction	.64	1	2	3	4	5
2.	The instructor has a sense of proportion and humour.	.57	1	2	3	4	5
3.	The instructor speaks clearly and distinctly.	.64	1	2	3	4	5
4.	The instructor speaks fluently and without hesitation.	.62	1	2	3	4	5
5.	The instructor selects materials suited to the needs of the class	.65	1	2	3	4	5
6.	The instructor expresses a willingness to help students.	.52	1	2	3	4	5
7.	The instructor is permissive and flexible.	.57	1	2	3	4	5
8.	The objectives of this course are clear	.60	1	2	3	4	5
9.	The instructor relates his subject to other fields.	.43	1	2	3	4	5
10.	The instructor exhibits self reliance and confidence.	.71	1	2	3	4	5
11.	Students feel free to express their own opinions.	.49	1	2	3	4	5

		item-subtest correlation					
12.	The instructor exhibits all around teaching ability.	.68	1	2	3	4	5
13.	The instructor is respectful of views other than his own.	.50	1	2	3	4	5
14.	The instructor presents the material in an interesting way.	.71	1	2	3	4	5
15.	The frequency of request for papers is adequate.	.48	1	2	3	4	5
16.	The instructor has a good knowledge of his subject matter.	.48	1	2	3	4	5
17.	The instructor stimulates students intellectual curiosity.	.73	1	2	3	4	5
18.	The instructor is knowledgeable of new developments in the field.	.65	1	2	3	4	5
19.	The frequency of exams is appropriate.	.47	1	2	3	4	5
20.	The course is well organized in meaningful sequence.	.69	1	2	3	4	5
21.	The course content contributes to general education.	.70	1	2	3	4	5
22.	Classes are well prepared.	.74	1	2	3	4	5
23.	Explanations are clearly presented.	.75	1	2	3	4	5
24.	The instructor is interested in his subject.	.71	1	2	3	4	5
25.	Materials are suitable to class level.	.71	1	2	3	4	5
26.	The instructors' style is entertaining.	.65	1	2	3	4	5
27.	The instructor is an effective teacher.	.61	1	2	3	4	5

Strongly Disagree 1 2 3 4 5 Strongly Agree

Hoyt Estimate of Reliability = .95
Standard Error of Measurement = 4.05

APPENDIX B

Study Habits Questionnaires

STUDY MOTIVATION SURVEY n = 82

NAME: _____

INSTRUCTIONS

Please circle the appropriate number beside each statement. Below is an explanation of the ratings:

1 = Never
2 = Not often
3 = Fairly often
4 = Frequent
5 = Very frequent

		item-subtest correlation					
1.	Do you often lose interest in your studies after the first few days or weeks?	.60	1	2	3	4	5
2.	Do you generally believe in doing on only enough to get a passing grade in your courses?	.35	1	2	3	4	5
3.	Do you frequently feel confused and undecided as to what your educational and vocational goals should be?	.27	1	2	3	4	5
4.	Do you believe that having a good time and getting one's full share of fun out of life is more important than studying?	.17	1	2	3	4	5
5.	Do you often spend the class period doodling or daydreaming instead of listening to the teacher?	.46	1	2	3	4	5
6.	Are you frequently unable to concentrate on your studies because of restlessness, moodiness, boredom, etc.?	.58	1	2	3	4	5
7.	Do you often feel that you are taking courses that are of little practical value to you?	.53	1	2	3	4	5

		item-subtest coorelation					
8.	Do you often feel like dropping out of school and getting a job?	.37	1	2	3	4	5
9.	Do you often feel that the things taught in school do not prepare on to meet adult problems?	.22	1	2	3	4	5
10.	Do you often dread reading test-books because they are so dull and boring?	.44	1	2	3	4	5
11.	Do you normally wait until a test is scheduled before reading testbook assignments or reviewing lecture notes?	.46	1	2	3	4	5
12.	Do you often feel that your teachers lack understanding of the needs and interests of students?	.20	1	2	3	4	5
13.	Do you normally hesitate to ask your teachers for help with troublesome assignments?	.24	1	2	3	4	5
14.	Do you generally feel reluctant to discuss future educational or vocational plans with your teachers?	.28	1	2	3	4	5

Hoyt Estimate of Reliability = 0.75

Standard Error of Measurement = 3.22

STUDY ORGANIZATION SURVEY

n = 82

NAME: _____

INSTRUCTIONS

Please circle the appropriate number beside each statement. Below is an explanation of the ratings:

1 = Never
2 = Not often
3 = Fairly often
4 = Frequent
5 = Very frequent

		item-subtest correlation					
1.	Do you usually put off preparing themes and reports until the last minute?	.52	1	2	3	4	5
2.	Do you often fail to complete homework assignments on time?	.30	1	2	3	4	5
3.	Do you frequently spend time reading magazines, watching television or exchanging gossip when you should be studying?	.46	1	2	3	4	5
4.	Do you usually wait a day or more before reviewing the notes taken in class?	.44	1	2	3	4	5
5.	Do you sometimes suddenly discover that an assignment is due sooner than you thought it was?	.44	1	2	3	4	5
6.	Do you often get behind in one course because of having to study for another?	.45	1	2	3	4	5
7.	Do you seem to accomplish very little in relation to the amount of time that you spend studying?	.41	1	2	3	4	5
8.	Is your study desk directly facing a window, door, or other source of distraction?	.12	1	2	3	4	5

		item-subtest correlation					
9.	Do you frequently have trouble organizing the content of a term paper or report?	.58	1	2	3	4	5
10.	Do you usually have trouble organizing the content of a term paper or report?	.52	1	2	3	4	5
11.	Do you sometimes prepare for a test by memorizing formulas, definitions or rules that you do not clearly understand?	.55	1	2	3	4	5
12.	Do you normally have difficulty organizing the material to be learned into logical study unit?	.48	1	2	3	4	5
13.	Do you depend mainly on last minute cramming in preparing for tests?	.46	1	2	3	4	5
14.	Do you frequently turn in your test paper without carefully rechecking for careless errors?	.38	1	2	3	4	5
15.	Are you frequently unable to finish answering essay questions within the allotted time?	.19	1	2	3	4	5
16.	Do you frequently lose points on ture-false tests because you failed to read the questions carefully?	.35	1	2	3	4	5
17.	Do you often spend too much time on the first half of a test and thus have to rush through the last half?	.34	1	2	3	4	5

Hoyt Estimate of Reliability = 0.85

Standard Error of Measurement = 3.31

STUDY TECHNIQUES SURVEY n = 82

NAME: _____

INSTRUCTIONS

Please circle the appropriate number beside each statement. Below is an explanation of the ratings:

1 - Never
2 = Not often
3 = Fairly often
4 = Frequent
5 = Very frequent

		item-subtest correlation					
1.	Do you frequently skip over the figures, graphs and tables in a reading assignment?	.42	1	2	3	4	5
2.	Do you frequantly have difficulty picking out the important points in a reading assignment?	.52	1	2	3	4	5
3.	Do you frequently catch yourself thinking about something totally unrelated to what you are reading?	.53	1	2	3	4	5
4.	Do you frequently have difficulty understanding your class notes when you try to read them over later?	.68	1	2	3	4	5
5.	Do you frequently get behind in your notetaking because you can't write fast enough?	.61	1	2	3	4	5
6.	Are your class notes usually a disorganized mess shortly after a semester begins?	.33	1	2	3	4	5
7.	Do you normally try to record your instructor's exact words when taking class notes?	.44	1	2	3	4	5
8.	Do you normally copy desired material word for word when taking reading notes?	.40	1	2	3	4	5

		item-subtest correlation					
9.	Do you usually keep photographs, mementos, or trophies on your study desk?	.17	1	2	3	4	5
10.	Is your study desk often so cluttered that you don't have enough room to study efficiently?	.42	1	2	3	4	5
11.	Do visitors to your room often interrupt your studying?	.39	1	2	3	4	5
12.	Do you often study with a television set, radio or phonograph playing?	.21	1	2	3	4	5
13.	Are magazines, pin-ups or hobby materials readily seen from where you usually study?	.33	1	2	3	4	5
14.	Is your studying often disturbed by activities and noises from outside your room?	.45	1	2	3	4	5

Hoyt Estimate of Reliability = .73

Standard Error of Measurement - 3.82