1992

Using Iva as a Predictor of Instructional Performance in a Graduate Management Program

Donald R. Carter
Loyola University Chicago

LOYOLA UNIVERSITY CHICAGO

USING IVA AS A PREDICTOR OF INSTRUCTIONAL PERFORMANCE IN A GRADUATE MANAGEMENT PROGRAM

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF CURRICULUM AND HUMAN RESOURCE DEVELOPMENT

BY

DONALD R. CARTER

CHICAGO, ILLINOIS

MAY 1992

ACKNOWLEDGEMENTS

I would like to acknowledge the many people who have made special contributions that led to the success of this research. First and foremost, I would like to thank my wife for her patience and encouragement during this research and for her shouldering the extra burden of dealing with our son and daughter while I was unavailable.

I am deeply indebted to Ms. Pat Surdyk for the hours spent keeping my parallelism and tenses on track during the writing of this dissertation. I would also like to thank my collegues at Keller Graduate School of Management for their assistance.

Last, but certainly not least, I would like to thank the members of my committee, Todd Hoover, Jack Kavanagh, Ron Morgan, and Patrick Mayers for their guidance and support. I owe a special debt of gratitude to my chairman, Todd Hoover, for creating IVA and for its many iterations over the course of this research. Also, to Patrick Mayers for teaching me what quality is all about.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

Chapter 1

INTRODUCTION

Ever since the publication of A Nation at Risk, there have been a myriad of programs suggested and implemented to improve education in this country. Ideas ranging from proficiency examinations in content areas for teachers to merit pay have been proposed, and in some cases implemented, as a means of improving instruction. The city of Chicago has attempted to effect major reforms by empowering local communities within the city to have a major voice in the governance of local schools. While the wave of change is sweeping the nations entire educational system, the change is taking different forms at different levels.

There are major differences in the emphasis given to reform between K-12 and higher education. For example, in higher education the emphasis seems to be in curriculum reform and not on improvements in teaching.[1] "So far, there has been almost no discussion in either the secondary or postsecondary communities about what individual teachers should be doing to improve learning in their own

---

[1]K. P. Cross, "The Adventures of Education in Wonderland: Implementing Education Reform" Phi Delta Kappan 68 no. 7 (1987): 496.

classrooms."[2] For the most part, the reform movement in higher education has been directed at using outcomes assessment in an attempt to measure student performance. Institutions of higher education, in response to pressure from state legislatures and accrediting associations are performing in-depth self-examinations to determine their effectiveness.

While it is difficult to find a commonly accepted definition of outcomes assessment, the following seems to capture the important attributes:

- "assessment tries to determine what students actually achieve in their college study; and
- assessment links educational objectives (of a course, a program, a field of study, or an institution) to some measure of student achievement."[3]

Hutchings and Marchese define assessment as a series of questions about student learning. The questions are as follows:

1. "What is the college's contribution to student learning? How and what do we know of that contribution?"

---

[2]Ibid.

[3]J. E. Rossmann and E. El-Khawas, "Thinking About Assessment: Perspectives for Presidents and Chief Academic Officers" (Washington, D.C.: American Council on Education and the American Association for Higher Education, June 1987), 1.

2. "Do our graduates know and can they do what our degrees imply? How do we ensure that?"

3. "What do the courses and instruction we provide add up to for students? Are they learning what we're teaching?"

4. "What knowledge and abilities do we intend that students acquire? Do they have opportunities to do so? Are they successful? At what level? Is that level good enough?"

5. "How can the quantity and quality of student learning be improved? What combination of college and student effort would it take to achieve higher levels of performance?"[4]

Without a clear-cut definition of what assessment should measure, colleges and universities have been left to their own devices to define their own assessment programs. The University of Virginia, under pressure from the State Council of Higher Education, endured several failed attempts to implement an outcoomes assessment program. The University's attitude and lack of direction caused conflict with the state agency. Only after assessment was viewed as a way to improve student performance and with faculty in control over the procedures did a program finally prosper.[5]

The University of Connecticut developed its assessment program through questions raised by the faculty who wanted to know if curriculum changes were working. Even with the support of faculty, the development of the program was

---

[4]P. Hutchings and T. Marchese, "Watching Assessment: Questions, Stories, Prospects" Change 22 (July/August 1990): 14.

[5]Ibid, 18.

achieved with difficulty. Lack of direction, some mistrust, and lack of properly defined goals all caused slow development of a school wide program. Even with considerable effort by the faculty there was still doubt as to the requirements mandated by the Department of Higher Education.[6] To them assessment, a worthwhile goal, was still an uncertainty.

At the other end of the spectrum, Alverno College shaped its entire program around outocomes assessment. Alverno's educational system revolves around a set of guiding principles that focus directly on the student and assessment. Assessment has become an integral part of the program at Alverno and this assessment repeatedly aids the college in developing its students according to its principles. "Assessment at Alverno focuses on the individual student. But to pursue the larger question about impact and effectiveness, the college employs sophisticated program evaluation. Alverno's office of research examines the impact, value, validity, and effectiveness of educational assumptions and programs, and work with faculty to refine the links between teaching, assessment practice, and long-term outcomes"[7]

It should be noted that, although outcomes assessment procedures are designed to look at student outcomes as a

---

[6]Ibid, 25.

[7]Ibid, 27.

measure of success, classroom performance by faculty is largely ignored. Tenure, academic freedom, and teacher contracts have come under fire as impediments to improvements in teaching performance. Tenure and contracts make it difficult for schools to evaluate and provide the staff development necessary to make meaningful improvements in classroom instruction. It is virtually impossible to dismiss an ineffective teacher, as the majority of administrators fear the almost certain resulting litigation. Academic freedom, the most cherished of all beliefs in colleges and universities, shares the blame for failure in ensuring that all students receive the same high level of instruction across sections and departments. Standardized terminal objectives, syllabi, and course rigor cannot be mandated across institutions.

With all of these criticisms aimed at the educational system, some ideas and methods have been forthcoming from business and industry in response to existing problems. Arthur Andersen and Company has entered into the politics of education with their new "School of the Future" program. This program proposes changes in teaching methods based on the concepts of "...simplify, automate, and integrate".[8] To implement this process, Arthur Andersen has developed a 14-point program, patterned after Deming's 14 points to

---

[8]R. L. Measelle and M. Egol, "A New System of Education: World-Class and Customer-Focused" (St. Charles, IL: Arthur Anderson Consulting, 1990), 2.

improving quality. The Anderson plan proposes some major changes to current teaching practices and the educational system and calls for a national commitment to educational improvements. While this program is specifically aimed at the K-12 system, there are points that appear have merit and applicability to the postsecondary system. A commitment to quality and a view of each student as a customer are two of the fourteen principles. A commitment to quality and ensuring customer satisfaction should become guiding principles of higher education, their sine qua non to improving education.

Quality

Business and industry view education differently than most colleges and universities, and they most assuredly view the meaning of a quality education differently. The meaning of the term quality in the educational community is vague in comparison to its meaning in business and industry. The concepts of total quality management date back to World War II. The Department of Defense believed these concepts to be so powerful that they were classified as secret. After the war, Dr. W. Edwards Deming, the father of quality, carried the concept to the Japanese, who unlike the Americans, accepted and implemented his teachings wholeheartedly. Quality, quality control, and total quality management are but a few of the terms describing quality. The Japanese Industrial Standards defines quality control as: "A system

of production methods which economically produces quality goods or services meeting the requirements of the consumer."[9] Juran defines quality as "fitness for use".[10] Most definitions of quality can be summarized as "meeting customer requirements--quality is measured by the degree of customer satisfaction with a product's characteristics and features."[11]

Students, parents, citizens, businesses, and the government of the United States are customers of the educational system. Given the definition of quality, is the present educational system meeting the demands of its customers? Some would say no. Given the diversity of the missions of the many institutions of higher learning, it would be inappropriate to make blanket statements of educational philosophy and try to apply the varied philosophies across all of the institutions, except in areas such as quality instruction. The definition of quality instruction for this study is customer satisfaction with "what" and "how" an instructor is teaching. Students are customers of education, the main purpose of education is to

---

[9]K. Ishikawa, What is Total Quality Control? The Japanese Way trans. David J. Lu (Englewood Cliffs, NJ: Prentice-Hall, Inc. 1985), 44.

[10]J. M. Juran and F. M. Gryna, Jr. Quality Planning and Analysis From Product Development Through Use 2nd Ed. (New York: McGraw-Hill, 1980), 1.

[11]The Ernst & Young Quality Improvement Consulting Group, Total Quality An Executives's Guide for the 1990s (Homewood, IL: Dow Jones-Irwin, 1990), 4.

meet their needs. Teachers, instructors and professors alike should be accountable for their performance in the classroom; they must meet the needs of their customers.

Faculty can be held responsible for their performance given that they are in a state of self-control. The concept of self-control provides a theoretical basis for the necessary change resulting from assessment. Juran's concept of self-control is described as: (1) knowing what to do, (2) knowing what is actually going on, and (3) taking regulatory action.[12] What makes Juran's concept of self-control important is that it allows for a separation of errors into two categories: (1) those that would be associated with the operator and, (2) those associated with management.[13] In trying to determine the primary cause for defects in the manufacturing of products, it was discovered that the majority of problems or errors was due to the failure of management to provide the necessary environment for workers to do a proper job. In a manufacturing setting, the knowledge of what to do can take many forms, from reviewing product samples to receiving verbal instructions from supervisors. The failure of an employee to understand a process will lead to quality failures. These quality failures are caused by poor communications within the system or by the improper design of products.

---

[12]J. M. Juran, Quality Planning, 314-323.

[13]Ibid.

For one to be in self-control, there needs to be a means of knowing whether one is performing to some standard. Employees such as machine operators must be able to measure performance during the manufacturing process. This feedback is used to help maintain the quality necessary during manufacturing. Feedback can be in many forms, but its purpose is to inform employees with respect to their performance.

Finally, the ability to regulate or make adjustments must be within one's capabilities. It is the responsibility of management to insure that any process can be changed and that it is capable of being changed.

Juran's concept can be applied to an educational setting. An instructor should be held responsible for his or her classroom performance if he or she is in a state of self-control. If not, the school, the school board, or state government needs to be held accountable for performance defects. The present educational system provides the knowledge of "supposed to do" by ensuring that all faculty are properly credentialed and are considered experts in the curricular area in which they will be teaching. Proper curriculum development, with valid goals and objectives, also need to be ensured. To provide the knowledge related to "is doing", faculty must have feedback mechanisms in place that provides a measures of effective performance in the classroom. Most of the feedback received needs to be

classroom based and related to outcomes assessment. Using the quality concept of continuous improvement, faculty can gather information on a weekly basis to provide the knowledge required to regulate performance. This performance regulation is totally under faculty control and covers such things as organization, student rapport, and providing feedback to the students. When these conditions for self-control are met, instructors can be held responsible for mandates that are properly funded and defined. Instructors can be held responsible for following curricula that are designed properly and in which they had input. Instructors can be held responsible for ensuring that classroom environments are conducive to learning. Instructors can be held responsible for poor performance when there is a commitment to provide the staff training necessary for improvement.

Terms such as effective performance have been extensively debated. For the research project to be described below, performance was defined as the score received on the global rating question of a student feedback form (Appendix A). "Effective performance" was defined as achieving a mean score of 3.0 on the overall rating question and "excellent instruction" was defined as receiving a mean score of 3.4 or better. Once again, excellent instruction is the goal to be sought in the classroom. This coincides

with one definition of quality education, that is, providing customer satisfaction.

Statement of the Problem

The overall purpose of this study was to investigate the usefulness of the Instructional Verbal Analysis (IVA) computer program as a means of predicting instructor performance in the classroom. IVA is based on the work of Ned Flanders whose original study was designed to provide feedback to teachers to assist them in becoming more effective.[14] Hoover[15] has modified Flanders' work for use with a computer system to gather data on verbal activity. IVA in its present form has been further modified to take into account not only verbal behavior but presentation behaviors as well.

Research Questions

The specific research questions to be addressed in the study were as follows:

1. Is IVA a useful predictor of faculty effectiveness as measured by the students' global rating?

---

[14]N. Flanders, Analyzing Teaching Behavior (Reading, Mass,: Addison-Wesley, 1970), 35.

[15]T. Hoover, "An Experimental Study of a Computer Assisted Teacher Training System Using Flanders' Interaction Analysis Category System Providing Immediate Feedback of Teaching Behavior to Naive Subjects" (Unpublished Doctoral Dissertation, University of Nebraska, 1975)

2. Are there differences in the presentation behaviors among the instructors?

Limitations

As with any field based study, there are limitations that must be considered when attempting to interpret the results. The limitations of this study are as follows:

1. The instructors evaluated for this study are not full-time faculty and had little training in educational theory.
2. The subject population was limited in number. There were only 70 courses taught within the geographic region during a term. Thirty percent of these instructors (21) participated.
3. The adult student population evaluating the faculty were all part-time and all enrolled in similar courses taught in an MBA program.
4. While randomization was used to select the instructors for the study, intact groups of students were used.

SUMMARY

The study was designed to investigate the usefulness of IVA as a predictor of teaching success. Teaching success for this study is defined as customer satisfaction with a particular course and instructor. This definition is based on the theories of total quality management by such people as Deming and Juran, noted experts in this field. The same

theories used by business can be applied to education as a basis for improvement. Juran's concept of self-control can be used as a guiding principle for holding instructors accountable for their performance in a classroom.

Chapter 2

LITERATURE REVIEW

Nothing seems to have caused more debate within college and university faculties than the use of student ratings. At one end of the spectrum are those who feel their use has caused an erosion of good teaching and scholarship[16] and at the other end are those who feel they are a legitimate and useful tool.[17] In a survey conducted by Marsh and Overall, faculty identified the following characteristics as causes of bias in student ratings: (1) course difficulty; (2) grading leniency; (3) instructor popularity; (4) student interest in the subject before taking the course; and, (5) students' GPA.[18] These are but a few of the biasing characteristics that faculty feel render student evaluations useless. Those who see the value of student evaluations

---

[16]R. B. Glassman, "Course Evaluations: Are Half of Us Really 'Below Average?'" Academe 74:44 (JL/AG 1988): 11.

[17]N. Tollefson, H. Wigington, and P. McKnight, "Course Ratings as Measures of Instructional Effectiveness" Instructional Science 12 (1983): 389.

[18]H. W. Marsh and J. U. Overall, "Validity of Students' Evaluations of Teaching: A Comparison with Instructor Self Evaluations by Teaching Assistants, Undergraduate Faculty, and Graduate Faculty," Paper presented at Annual Meeting of the American Education Research Association, San Francisco, 1979, ERIC Document Reproduction Service No. ED177205.

contend that the ratings do reflect differences in faculty effectiveness and that the biases are overstated. In cases where research on biasing characteristics produced statistically significant relationships, that significance is minimal at best.[19]

In this Chapter a review of the literature relevant to this research will be presented. There is an extensive body of knowledge in existence concerning the use of student/faculty evaluations. The information will be divided into the following subheadings:

- characteristics of teaching effectiveness
- reliability and validity of evaluations
- biasing characteristics

An additional review of the literature will focus on adult learning.

Characteristics of Teaching Effectiveness

One of the key factors in the effort to improve instruction has been the attempt to identify the qualities of an excellent instructor as perceived by students. A meta-analysis in 1976 by Feldmen analyzed the research on students' views of effective teaching. The studies that he analyzed gathered student opinions in four ways:

---

[19]P. C. Abrami and D. A. Mizener, "Does the Attitude Similarity of College Professors and Their Students Produce 'Bias' in Course Evaluations?" American Educational Research Journal, 20 (1983):123-136.

- characteristics that students reported as being most associated with ideal or best teachers and as most important to effective teaching, with students furnishing lists of characteristics of their own choosing

- characteristics that students report as being most associated with ideal or best teachers and as most important to effective teaching, with students responding to pre-set lists of characteristics

- specific items on teacher evaluation questionnaires that are most strongly associated with the global evaluation of the instructor

- specific evaluation items that most frequently combine with global evaluation items to form the highest loadings on the same factor in factor-analytic studies.

The following, based on the review of research, are the characteristics Feldman identified as important to successful teaching:

1. Instructor stimulated interest.
2. Instructor's enthusiasm for the subject or teaching.
3. Instructor's knowledge of the subject matter.
4. Instructor's intellectual expansiveness and intelligence.
5. Instructor's preparation and organization of the course.
6. Clarity and understandableness.
7. Instructor's elocutionary skills.
8. Sensitivity to, and concern with, class level and progress.
9. Clarity of course objectives and requirements.
10. Nature and value of the course materials including their usefulness and relevance.

11. Nature and usefulness of supplementary materials and teaching aids.

12. Difficulty and workload of the course.

13. Instructor's fairness and impartiality of evaluation of students; quality of exams.

14. Classroom management.

15. Nature, quality and frequency of feedback from instructor to student.

16. Instructor's encouragement of questions and discussion, and openness to the opinions of others.

17. Intellectual challenge and encouragement of independent thought.

18. Instructor's concern and respect for students; friendliness of the instructor.

19. Instructor's availability and helpfulness.[20]

Feldman concluded that there were certain factors that were consistently associated with effective instruction across all methods of data collection: stimulation of interest; clarity and understandableness; knowledge of subject matter; preparation for, and organization of, the course; and enthusiasm for the subject matter and for teaching. Also, students stressed the need for an instructor to be friendly, helpful, open to other opinions, and available.[21]

Pohlmann analyzed approximately 30,000 student evaluations to determine effective instruction across five different disciplines i.e., science and mathematics,

---

[20]K. A. Feldman, "The Superior College Teacher From the Students' View Research in Higher Education 5 (1976): 243.

[21]Ibid., 243.

education, social sciences, humanities, and business. Although there were differences between disciplines, four factors correlated highly across all of them to characterize effective instruction: (1) the instructor knew if students understood him or her; (2) the instructor answered impromptu questions satisfactorily; (3) the instructor achieved the specified objectives of the course; and, (4) the instructor gave several examples to explain complex topics.[22]

In a study by Truex, using Flanagan's critical incident technique, several factors critical to teaching performance were identified. Flanagan's technique involves an observer recording extremes of behavior, or critical incidents, during a classroom visitation. The result of this technique is "... a derived and reliable statement of the facets of performance which are crucial to success or failure in teaching performance."[23] Truex classified the observations under personal/social and professional, each with their own subgroupings. During the evaluation it was ascertained that personal/social factors were of lesser importance than the professional at the college level. The following factors were identified as crucial for effective college level teaching: (1) knowledge of subject matter; (2) class

---

[22]J. T. Pohlmann, "A Description of Effective College Teaching in Five Disciplines as Measured by Student Ratings" Research in Higher Education 4 (1976): 335.

[23]M. H. Truex, "Factors Critical to College Teaching Success or Failure" Improving College and University Teaching 23 no. 4 (1975): 236.

presentation; (3) ability of the instructor to evoke meaningful classroom response; (4) enthusiasm for subject and learning; and, (5) empathic understanding of the student. In two other studies that used Flanagan's technique, one by Menges and Kulieke[24] and another by Ross[25], clarity of lecture and the ability to elicit student involvement, and planning and organization of lectures were the top three factors rated by students as critical to their learning. The negative experiences related by students focused on inconsistencies in lectures, confusing explanations and disorganization by the instructor.[26]

Not all of the studies of effective teaching have been quantitative in nature. Guskey[27] used in depth clinical interviews of highly effective instructors as identified by both students and academic deans. In analyzing the data for the interviews Guskey found that there were no commonalities in personal characteristics among the instructors interviewed, except that nearly all had teaching certificates and had some formal training in education in

---

[24]R. J. Menges and M. J. Kulieke, "Satisfaction and Dissatisfaction in the College Classroom" Higher Education 13 (1984): 255.

[25]J. M. Ross, "Critical Teaching Behaviors as Perceived by Adult Undergraduates" Paper presented at the Annual Meeting of the American Education Research Association, March 27-31, 1989, 17, EDRS ED 311015, microfiche.

[26]Ibid.

[27]T. R. Guskey, Improving Student Learning in College Classrooms (Springfield, IL: Charles C. Thomas, 1984), 15.

addition to particular disciplines studies. This lack of common personal characteristics is highly contrasted with the commonalities in teaching behaviors. Guskey classified the teaching behaviors into four categories: (1) planning, organization and cues; (2) positive regard for students; (3) student participation; and (4) feedback, correctives and reinforcement.

Planning and organization take a considerable amount of the effective teacher's time. Course outlines and syllabi that include descriptions, objectives and grading criteria as well as actual assignments are given to the students at the beginning of the course. As the course continues, each lesson is clearly planned and organized with a clear structure and format. Even with highly structured class meetings, flexibility and concern for students is always evident. During class sessions, effective instructors constantly probed for student understanding and provided plenty of examples and illustrations of practical applications to reinforce the concepts being taught. These instructors also stressed being active during each class session. They would frequently move about the classroom and ask questions of their students; class participation is a major guiding principle. Finally, these effective instructors provided regular and specific feedback to their students. Feedback was provided either through comments on exams and assignments or in individual conferences.

Feedback in the form of praise was also considered extremely important.[28]

Reliability of Evaluations

According to Kerlinger, the definition of reliability is the accuracy or precision of a measuring instrument.[29] In conjunction with student ratings, Marsh defines reliability as the relative lack of random error.[30] Several techniques can be used to determine reliability of a student rating instrument. A reliable item, also known as interclass correlation, is one in which there is agreement among ratings within each class, but differences between different classes. A split-half form is another method used to determine an instrument's reliability. Using this method, random halves of a ratings form are taken from each of a large number of classes and then are correlated with one another. Cronbach's coefficient alpha is another method of determining reliability. Coefficient alpha differs from interclass correlation in that it does not include

---

[28]Ibid., 15-27.

[29]F. N. Kerlinger, Foundations of Behavioral Research 3d ed., (New York: Holt, Rinehart, and Winston, 1986), 405.

[30]H. W. Marsh, "Students' Evaluation of College/University Teaching: A Description of Research and an Instrument" (Presented at the Annual Meeting of the Australian Association of Research in Education, Sydney, Australia, November 1980): 7.

disagreement among students within the same class as a source of unreliability.[31]

According to Remmers[32] the correlation between any two students in the same class is small but the reliability of the class average response is good. For example the reliability of the Students' Evaluations of Educational Quality (SEEQ) evaluation instrument for a class of 25 is 0.90 but for classes with 10 or less the reliability rating is about 0.20.[33] In an investigation, Doyle reported a reliability range of .80 to .89 for the student ratings.[34] In the same study, ratings for colleagues was less than those of the students, ranging from .65 to .86.[35]

Validity of Evaluations

A large number of studies deal with the subject of validity of student evaluations. The trend has been to establish a link between student ratings and other measures of teaching effectiveness. Other measures which researchers

---

[31]Ibid.

[32]H. H. Remmers, "Reliability and Halo Effect on High School and College Students' Judgements of Their Teachers" Journal of Applied Psychology 18 (1934): 620.

[33]H. W. Marsh, "Students' Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research," International Journal of Educational Research 11 no. 3 (1987): 275.

[34]K. Doyle, Student Evaluation of Instruction (Lexington, Massachusetts: Lexington Books 1975), 19.

[35]Ibid.

have used are student test scores, ratings by peers, ratings by administrators, ratings by alumni, and self assessments. There have been many studies supporting the validity of student ratings and some that discount this premise. Gaski[36] presents a summary of some of the studies in Figure 1.

Gessner (1973)	78 students	High correlation between evaluation and performance.
Frey (1973)	13 instructors, 354 students	Strong relation between ratings and teaching quality.
Marsh, Fleiner, & Thomas	18 sections, 720 students	Student evaluations positively correlated with final exam and scores predicted by SAT profile.
Marsh (1977)	62 instructors, 591 classes, 1847 students	Evaluations validated with retrospective reports of most/least outstanding.
Marsh, Overall, and Kesler (1979)	51 instructors, 83 courses	Factor analysis indicated similar student-faculty evaluation dimensions across evaluation factors; higher student ratings courses instructor rated most effective.
Marsh & Overall (1980)	31 sections, approximately 960 students	Generally and moderately positive relations between student ratings and teaching effectiveness criteria.

[36]J. F. Gaski, "Construct Validity of Measures of College Teaching Effectiveness" Journal of Educational Psychology 79, no. 3 (1987): 327.

Howard & Maxwell (1980)	Two experiments	Weak, positive relation between expected grades and student satisfaction; student motivation and performance explained more of variation in satisfaction.
Marsh (1982)	329 classes	General agreement between student and instructor ratings in multitrait-multimethod analysis.
Howard, Conway, & Maxwell (1985)	43 instructors, 34 students, 30 former students	Student and former student ratings reported superior in convergent/discriminant validity to other methods; that is, self, colleague, and trainer observer ratings.
Rodin & Rodin (1972)	293 students	Inverse partial correlation between objective measure of amount learned and student rating.
Snyder & Clair (1976)	72 students	Expected grades inversely related to evaluations; perceived obtained grades positively related.
Pratt & Pratt	175 students	Very little correlation between obtained grades and student ratings; strong positive correlation between expected grades and ratings.
Brown (1976)	2,360 sections, 30,000 students ratings	In stepwise regression, grades represent a more powerful predictor of ratings than any other hypothesized antecedent.
Powell (1977)	5 sections, 35-45 students per section	Ratings of instructor decreased as grading stringency increased; amount learned increased as grading stringency increased.

Biasing Characteristics

There are many who are against the use of student evaluations and have stated that there are a number of influences on the students which influence validity. These other influences or biases have been investigated thoroughly with mixed results.

One of the most recent and interesting studies by Tollefson, Chen, and Kleinsasser investigated student/teacher attitudes as a biasing factor. The researchers were operating under the premise that those with similar attitudes are attracted to each other causing a higher evaluation. Findings indicated that it was the differences in teachers not the similarity of attitudes between instructor and students that affect student ratings.[37]

Personality of instructors has also been examined as a biasing factor. In his research, Jones examined whether students are able to separate the instructor's personality from their perceptions of effective instructors. Jones used an approach described by Scriven[38] in attempting to separate "irrelevant" context variables that might distort a

---

[37]N. Tollefson, J. S. Chen, and Kleinsasser, "The Relationship of Students' Attitudes About Effective Teaching" Educational and Psychological Measurement 49 (1989): 529.

[38]M. Scriven, "Summative Teacher Evaluation" in J. Millman (ed.) Handbook of Teacher Evaluation (Beverly Hills: Sage Publications, 1981), 244, quoted in J. Jones, "Students' Ratings of Teacher Personality and Teaching Competence" Higher Education 18 (1989): 552.

student's rating. Scriven suggests that evaluations should specifically ask for an expression of liking for the instructor as a person and then the subject matter. After this is done the students would then be asked to evaluate the instructor's job in teaching the course. The results of Jones' research indicates that personality, as a whole, is related to the students' perception of the instructor's ability to teach.[39]

Several researchers have studied specific personality traits and their relationships to student ratings. In a recently completed study Murray, Rushton, and Paunonen investigated the effects of 29 personality traits on student ratings of teaching effectiveness. Forty six psychology teachers were evaluated over a period of time. The study examined peer ratings (ratings of other faculty and administrators) and student ratings of these faculty across the same and different courses taught. Three important findings were ascertained from this research. There was evidence that most instructors receive a wide range of ratings across different courses they have taught, while the ratings are fairly stable over time for the same course. The second finding was a strong relationship between peer and student ratings. The evaluations of faculty by other faculty or administrators are highly correlated with those

---

[39]J. Jones, "Students' Ratings of Teacher Personality and Teaching Competence" Higher Education 18 no. 5 (1989): 551.

evaluations of students. Finally, leadership, extraversion, liberalism, supportingness, intellectual curiosity, and changeableness were the personality traits that correlated the highest with the teaching effectiveness ratings. In addition, specific personality traits were correlated differently for different types of courses. Also, for graduate and senior honors courses, instructor achievement and endurance were the most significant traits; these traits were unrelated to any other undergraduate course.[40] Abrami and others investigated the personality factor from the student perspective. In a series of four studies, Abrami attempted to discover if the attitudes, traits, and values of the students would affect the faculty ratings. It was concluded that student personality has no effect on faculty ratings. They did support the finding that instructor personality does affect teacher effectiveness ratings.[41]

Another area of research closely related to that of personality is that of instructor expressiveness. The "Dr.

---

[40]H. G. Murray, J. P. Rushton, and S. V. Paunonen, "Teacher Personality Traits and Student Instructional Ratings in Six Types of University Courses" Journal Educational Psychology 82 no. 2 (1990): 250.

[41]P. C. Abrami, R. P. Perry, and L. Leventhal, "The Relationship Between Student Personality Characteristics, Teacher Ratings, and Student Achievement" Journal of Educational Psychology 74 no.1 (1982): 111.

Fox" study by Naftulin[42] was a controversial study often cited by critics questioning the usefulness of student evaluations. In this research, Naftulin used an actor to play the part of an instructor (Dr. Fox) and lecture to a group of educators at a national conference. Upon completion of the lecture the group rated the performance of the lecturer as effective, even though the lecture was devoid of substance. Even with its weak methodology the research is used by critics as an example of how expressive faculty can seduce students into giving higher evaluations than actually deserved.

There is no doubt that teacher expressiveness is an important characteristic of effective instruction, but it is not the sole determinant that students will use in evaluation of instruction. Perry, Marsh and Ware, and Abrami, in a meta-analysis, have concluded that educational seduction is not supported by existing research.[43] In a study by Perry and others, an attempt was made to replicate a study by Williams and Ware[44] which found that "... teacher differences in expressiveness controlled the degree to which

---

[42]D. H. Naftulin, J. E. Ware, and F. A. Donnelly, "The Doctor Fox Lecture: A Paradigm of Educational Seduction" Journal of Medical Education 78 (1973): 630.

[43]Marsh, Students' Evaluation of Teaching, 333.

[44]R. G. Williams and J. E. Ware, "Validity of Student Ratings of Instruction Under Different Incentive Conditions: A Further Study of the Dr. Fox Effect" Journal of Educational Psychology 68 (1976): 48.

lecture content affected student ratings differently from student achievement."[45] The Perry study failed to replicate the findings of the other study.

Grades, class size, faculty rank, and gender, have also been researched as biasing factors of faculty ratings. The research in these areas have mixed results. Results indicate that none of the factors have a significant effect on student ratings of the faculty and should not be used as evidence to question the validity of student ratings.

Adult Learning

The student population for this research was all adults. At most colleges and universities today, the adult student population comprises more and more of the total population. For evaluation purposes it is necessary to know if the effective teaching research is appropriate to the adult learner.

Program development for adults should be guided by the appropriate principles. Brookfield has identified six such principles of effective practice. The following principles should be used as guidelines for curriculum development.

- "Participation in learning is voluntary; adults engage in learning as a result of their own volition. It may be that the circumstances prompting this learning are external to the learner, but the decision to learn is the learner's. Hence, excluded

---

[45]R. P. Perry, P. C. Abrami, and L. Leventhal, "Educational Seduction: The Effect of Instructor Expressiveness and Lecture Content on Student Ratings and Achievement" Journal of Educational Psychology 71 no. 1 (1979): 107.

are those settings in which adults are coerced, bullied, or intimidated into learning.

- Effective practice is characterized by a respect among participants for each other's self-worth. Foreign to facilitation are behaviors, practices, or statements that belittle others or that involve emotional or physical abuse. This does not mean that criticism should be absent from educational encounters.

- Facilitation is collaborative. Facilitators and learners are engaged in a cooperative enterprise in which, at different times and for different purposes, leadership and facilitation roles will be assumed by different group members.

- Praxis is placed at the heart of effective facilitation. Learners and facilitators are involved in a continual process of activity, reflection upon activity, collaborative analysis of activity, new activity, further reflection and collaborative analysis, and so on.

- Facilitation aims to foster in adults a spirit of critical reflection. Through educational encounters, learners come to appreciate that values, beliefs, behaviors, and ideologies are culturally transmitted and that they are provisional and relative. This awareness that the supported givens of work conduct, relationships, and political allegiances are, in fact, culturally constructed means that adults will come to question many aspects of their professional, personal, and political lives.

- The aim of facilitation is the nurturing of self-directed, empowered adults."[46]

There are many studies that try to identify the principles of how adults learn. Most of the research is qualitative in nature, based on conversations and observation. Brookfield summarizes the work of many of these researchers (Gibb, Miller, Kidd, Knox, Brundageand and

---

[46]S. D. Brookfield, Understanding and Facilitating Adult Learning (San Francisco: Jossey-Bass Publishers, 1986), 9.

Mackeracher, Smith, and Darkenwald and Merriam) to formulate a list of guiding principles that should assist the curriculum development for the adult learner. Adults learn their entire lives; they have different learning styles; they like their learning activities to be problem based with practical and immediate applications. Also, past experiences will have a profound effect on their learning; they need to have early successes, and they show a tendency to prefer self-directedness.[47]

For any program to be successful, it must have the full support of its constituents. In today's climate of outcomes assessment to insure programs are fulfilling their stated purposes, it is important to understand the motivations and concerns of the students involved in any given program. In a study by Pierson and Springer, adults in the program indicated that they felt they were independent and self-motivated learners. The adults also felt less comfortable with their academic skills, especially math and writing. When asked to identify the reasons for being in school, these adults indicated that in addition to personal satisfaction and development, increase in income, development of new career potential, and increase in specific job skills were most important.[48]

---

[47]Ibid., 31.

[48]M. J. Pierson and S. B. Springer, "Can Anything Good Come from Non-traditional Degree Programs?" Lifelong Learning: An Omnibus of Practice and Research 11 no. 5 (1988): 20.

In her study of adults, Handler investigated the effects of math anxiety on the performance. Her study shows that anxiety affects both men and women and it results in a high level of emotional interference that can disrupt memory and logic.[49]

Loesch and Foley investigated learning styles to see if there were any differences in learning styles among adults enrolled in both traditional and nontraditional programs.[50] They discovered that students in nontraditional programs preferred to organize their own programs of study while those in more traditional studies preferred a more structured teacher directed environment. Ostmoe et al., in a similar study found that nursing students (traditional environment) preferred a highly structured and organized teacher-directed environment.[51]

Much of the adult learning research has focused on adults as a single group, many researchers are now investigating gender to see if differences exist in adult learning. One such area is the study of classroom social

---

[49]J. Handler, "Math Anxiety in Adult Learning" Adult Learning (April, 1990): 20-21.

[50]T. Loesch and R. Foley, "Learning Preference Differences Among Adults in Traditional and Nontraditional Baccalaureate Programs" Adult Education Quarterly 38 no. 4 (Summer, 1988): 224.

[51]P. Ostmoe, H. Van Hoozer, A. Scheffel and C. Crovell, "Learning Style Preferences and Selection of Learning Strategies: Considerations and Implications for Nurse Educators" 23 no. 1 (1984): 27.

climate. With reentry of significant numbers of women returning to higher education, investigation of these differences is becoming more prominent. Gilligan's research indicated that women and men differ in their sense of self. Men link their accomplishments with a description of themselves, while woman develop a sense of self that is developed around relationships.[52] Beer and Darkenwald investigated the differences between men's and women's perceptions of the social environments of a classroom. Their research indicates that there are differences between the sexes and that women perceive more affiliation and a greater degree of involvement in the classroom than do men.[53] This research will lead instructors to change teaching styles to be more responsive to the needs of returning female students.

Summary

Even though there is not complete agreement, certain characteristics of effective teaching have been identified that affect the ratings of students. There are five characteristics that appear to be especially influential in determining student satisfaction with a particular

---

[52]C. Gilligan, In a Different Voice (Cambridge, MA: Harvard University Press, 1982), 23.

[53]C. T. Beer and G. G. Dardenwald, "Gender Differences in Adult Student Perception of College Classroom Social Environments" Adult Education Quarterly 40 no. 1 (Fall 1989): 40.

instructor: enthusiasm, organization, rapport, student participation, and feedback. These factors will be examined in this research.

The research on student ratings of faculty is one of the most widely studied areas, and the body of knowledge continues to grow. Based on the material presented, it is evident that faculty evaluations are both a reliable and valid tool to use in making decisions regarding teaching performance in the classroom. Research has been conducted on many facets of student ratings, including many sources of potential bias, that some believe reduces the validity of student ratings. Personality, attitudes, effective teaching characteristics, class size, gender, and faculty status have all been investigated as possible biasing characteristics. However, a careful analysis of the research has shown that even when one of these factor was shown to have an effect, the size of the effect was too small to be significant.

Regardless of how those opposed to student ratings may feel, student ratings will continue to be used as a measure of faculty performance in the classroom. Student ratings will have an effect on personnel decisions and tenure as teaching performance becomes more highly regarded as a function of higher education across all types of institutions.

While research in adult education is quite broad, there is considerable agreement on guiding principles for adult

education. Adults are learning of their own volition, they need early successes in their education to provide the motivation to continue. Adults prefer a cooperative environment, and male and female students have different needs in the classroom. Educators need to meet the needs of this growing segment of the student population to ensure their education is both rewarding and successful.

Chapter 3

RESEARCH PROCEDURES

Many studies have been designed in an attempt to identify the essential characteristics of good teaching or what distinguishes a successful teacher from those who are not. None of the studies reported in the literature have been designed to predict teaching success by measuring in class behaviors of effective teaching faculty. An effort was made in the study described here to build and test a model designed to predict successful teaching performance as measured by students as an indication of customer satisfaction. In this chapter, the methodology for this study is discussed. A description of the subjects is given, field procedures are reviewed, methods used for data collection are summarized, and the general research design is explained.

Subjects

The subjects for this study were faculty (n=24) from a practitioner-oriented graduate school of management. The school is a publicly held company that offers an MBA degree and a highly specialized masters degree in project management. The school has a unique philosophy when

compared to traditional graduate schools of management; its mission is to provide quality practitioner oriented graduate management education through excellence in teaching while meeting the needs of its working adult population. In meeting its mission, the school has decentralized its delivery systems within its intrastate operating region and has expanded to other states. The faculty are part-time; most are middle- and upper-management business practitioners, and all have at least 10 years of business experience with at least 5 years in the content area for which they are teaching. They also have a Master of Business Administration (MBA) degree or its equivalent. The subjects were selected at random from a pool representing faculty from the school's five Illinois locations. The group was divided into two subgroupings representing faculty who are experienced and have met minimum performance standards by receiving at least a 3.0 (on a 4 point scale on which Excellent = 4, Good = 3, Fair =2 and Poor = 1) rating on the student feedback form and new faculty who were not yet rated. According to school policy, faculty who do not attain minimum performance levels are not invited to teach again unless the Associate Academic Dean believes that improvement can be made, and the individual faculty member has made a commitment to improve. A major goal of the school is to have faculty who consistently achieve at a 3.4 rating or better on the 4.0 scale. It should be noted that

the student feedback form, the instrument used for faculty evaluation, has one global rating question.

Description of the Instructional Environment

Each instructor who was evaluated in this study was a practicing business professional teaching part-time. Before being hired, a potential instructor had to express a desire to undergo an extensive required training period. During the hiring process, the applicant was interviewed twice and, during the second interview, provided a 15 minute presentation to demonstrate his/her ability to perform in front of a small group. After the hiring decision had been made, there were three group training sessions conducted by the Associate Academic Dean (AAD) and/or Center Director (CD) in which discussions took place related to such matters as teaching techniques, lesson planning, test preparation, grading, etc. A teaching model was presented that had proven to be successful for the instructors in the program. Class time was organized to provide continuity from week-to-week. Each session began with a discussion of homework, new material was presented, examples were given that demonstrated the concepts being taught, students worked though problems under the instructor's guidance and homework problems were assigned. The performance ratings of those instructors who followed this model indicated that this week-to-week consistency was preferred by the adult students enrolled in the program.

Each new instructor and continuing faculty doing new course preparation was required to prepare extensively before the first class meeting. During the three training sessions, the AAD provided new instructors with a curriculum guide for the course being taught and a copy of the faculty handbook. The curriculum guide provided the terminal course objectives, topic outline, sample syllabus, midterm and final exam. It also contained suggested homework assignments and suggestions on week-to-week organizational flow. The new instructor was required to meet with a veteran instructor and visit a class. For the second training session, the new instructor had to provide a proposed rough draft of the syllabus, which was reviewed by the appropriate curriculum coordinator, and was required to give a 20 minute presentation covering a topic scheduled for the first night of class.

The curriculum coordinators (CC) served as content experts within the school; their function was to provide leadership in curriculum development. With the assistance of existing faculty, the curriculum coordinators insured that the curriculum stayed current and maintained its practitioner focus. The CCs systematically reviewed the materials of a new instructor, analyzed these materials to insure that terminal objectives were covered and measured by either an exam, project, or some other means.

During the instructor's 20 minute presentation, the audience role-played to provide a realistic simulation of an actual class situation. A critique of the teaching demonstration was done using the teaching model described earlier, as a guideline for performance.

A final training session is scheduled to assist the new instructor in final preparation of lesson plans, class notes, and midterm and final examinations. The new instructor was encouraged to review supplemental materials of other instructors and to incorporate weekly readings to supplement the text. These readings were supplied to the students when textbooks were purchased, and were used by the Associate Academic Dean as a basis of providing additional help to the instructor.

Continuing faculty have individual development plans that were prepared and reviewed on a continuing basis by the Center Director with assistance from the curriculum coordinators. Their syllabus and exams were reviewed regularly by the curriculum coordinators, any concerns resulting from these reviews were conveyed to the CDs who then took what they considered to be appropriate action.

A new instructor was evaluated twice during the 10-week term by the AAD and/or the CD. Continuing faculty were evaluated once during the term by the CD. Meetings were subsequently held with the instructor to discuss the observations and to offer any suggestions to help improve

instruction. After the midterm and final examinations were administered, a meeting was scheduled to discuss grading procedures and the assignment of grades before the students received the results of their examinations. Finally, during the ninth week, the students rated the instructor. One last meeting then occured to discuss the evaluations, any improvements that had been made or that were to be made, and future teaching possibilities.

Students

The student population is composed of adult learners, with a mean age of 32, who had been in the work force an average of ten years prior to their decision to pursue an MBA. Approximately 56% of the student population did not have a business-related undergraduate degree and a significant number were employed in occupations outside traditional business-related fields such as nursing or teaching. The majority of the students, 71%, were male, approximately 29% were female, and 10% were minorities. Only 11% could be classified as full-time students. Table 3-2 presents a breakdown of undergraduate majors, extracted from individual students' transcripts. Table 3-3 presents the professional occupations of the students as described on their admissions applications.

Instrumentation

Each of the participants was evaluated once during the

term in either the second, fourth, seventh or eighth week of

Table 3-2.--Undergraduate Majors

Business Administration	39%
Engineering	15%
Science/Math	11%
Social Sciences	10%
Humanities/Art/Philosophy	9%
Computer Science	7%
Economics	5%
Education	4%

Table 3-3.--Student Occupations

Marketing and Sales	15%
Engineering	10%
Finance	9%
Accounting	7%
Data Processing	7%
Health Care Related	7%
Manufacturing Related	5%
Human Resources	3%
Quality Control	2%
Self-Employed	2%
Other	33%

instruction. The evaluator used the Instructional Verbal Analysis System (IVA) and a laptop computer to gather data.[54] IVA, a computer program, was designed to function as a self-assessment tool for student teachers. Student teachers were observed during a student teaching session by an observer using IVA to record the verbal interchange between the student teacher and the students. At a later

[54]T. Hoover, "Guidebook and Directions for a Computer Program Titled IVA," (Unpublished Work in Progress, Loyola University of Chicago, 1989), 3.

date, the student teacher used the IVA results with an analysis form to perform a self-assessment. Using the results of the analysis the student teachers, then embarked on a program to modify their teaching behavior to conform more to Flanders' theories and improve their instructional techniques.

As noted earlier IVA was based on the work of Ned Flanders, who during the 1960s developed the Flanders Interaction Analysis Category system (FIAC).[55] FIAC was designed to be used to provide feedback to teachers by assisting them in changing behaviors for more effective teaching. Flanders' system was a manual system which relied on a matrix for interpretation. Many researchers have used modified versions of Flanders' work with varying results. Flanders published his most extensive analysis of FIAC in 1970 in which he discussed the results of these studies. He stated that the "ultimate goal [of FIAC] is to explain and predict the consequences of different patterns of interaction, strung together into sequences which can be called teaching strategies."[56] In his analysis of the studies based on his work, he concluded "that when classroom interaction patterns indicate that pupils have opportunities to express their ideas, and when these ideas are

---

[55]N. A. Flanders, Analyzing Teaching Behavior (Reading, MA: Addison-Wesley, 1970), 35.

[56]N. A. Flanders, Analyzing Teaching Behavior (Reading MA: Addison-Wesley, 1970), 400.

incorporated into the learning activities, then the pupils seem to learn more and to develop more positive attitudes toward the teacher and the learning activities."[57] Hoover was the first to use Flanders' ideas with a computer system.[58] He subsequently developed IVA with further modification to the original Flanders system.

Using IVA, a classroom observer enters a code that classifies the verbal activity taking place at a particular moment. A code is entered every three seconds for the entire length of the observation, usually 30 minutes. Every four minutes IVA switches to an alternate screen where five additional factors of instruction are measured on a scale from one to five, with one being high and five being low.

There are ten categories that describe the verbal activities that take place in a classroom (Table 3-4). The ten categories that measure verbal activity are taken from Hoover's research. The alternate screen categories were added to IVA based on the results of the extensive research on faculty evaluation and characteristics of successful instruction (Table 3-5). These categories also represent types of communication taking place in an

---

[57]Ibid., 401.

[58]T. Hoover, "An Experimental Study of a Computer Assisted Teacher Training System Using Flander' Interaction analysis Category System Providing Immediate Feedback of Teaching Behavior to Naive Subjects" (Unpublished Doctoral Dissertation, University of Nebraska, 1975)

Table 3-4.--IVA Screen One

1. Clarify/Answer Questions;
2. Praises or Encourages;
3. Accepts/Uses Ideas of Learners;
4. Asks Questions;
5. Lecturing/Gives Information;
6. Gives Directions/Organizes;
7. Learner responds to a specific question;
8. Learner Initiates Own Comment or Responses;
9. Learner Asks Questions;
0. Silence or Confusion.

instruction session. The first categories are classified into 3 areas: (1) instructor-initiated communication; (2) student-initiated communication; or (3) no communication. The second group of categories represent instructor initiated communication factors that have been identified as representing effective instruction.

Upon the completion of the observation, the frequencies for each of the categories in group one are tallied by the computer and four ratios are calculated. A printout of the session can be furnished on request (Table 3-6). These ratios characterize the instructional behaviors exhibited during the class session.

The frequencies are used to form a matrix from which the ratios are calculated (Figure 3-7). It can be used for pattern analysis as described by Flanders in his original works. The matrix is a 10 x 10 table that corresponds to the categories used by the observer to record the teaching behaviors. Each pair of behaviors starting with X_1 (the first recorded behavior) and X_2 (the second recorded

Table 3-5.--IVA Screen Two

Instructor Enthusiasm Descriptors	Instructor Clarity Descriptors
Speaks expressively	Uses Examples
Moves while lecture	Multiple Examples
Gestures with hands	Practical Application
Facial Expression	Stresses points
Uses Humor	Repeats Ideas
Task Orientation Descriptors	**Instructor Rapport Descriptors**
Advise on exams	Friendly
Provides sample tests	Shows concern
Proceeds rapid pace	Offers help
States objectives	Tolerant
Instructor Organization Descriptors	
Outline on board/Overhead	
Gives overview of lecture	
Signal topic transition	
Explain how topic fits in	

behavior) is inserted into the matrix by the row and column designator (Mx_1,x_2) corresponding to the actual code number. For example, if the first two behaviors were 1 and 5, a count of 1 would be placed in cell 1,5. The third recorded behavior is then paired with the second behavior and the count in that cell is increased by one. This sequence of using the second observation of the previous pair with the next unrecorded observation to form the cell address continues until all of the data is recorded.

Table 3-6.--Sample of Analysis Output

The Instructional Verbal Analysis System
RATIO DISPLAY
Donald R. Carter

	ACHIEVED
THE RESPONSIVE RATIO IS:	5
THE DOMINANT RATIO IS:	94
THE QUESTIONING RATIO IS:	31
THE INITIATIVE RATIO IS:	71

Category Name	# Entries	Percentage
#1 Instructor Answers Question:	0	0%
#2 Instructor Praises:	0	0%
#3 Instructor Uses Ideas:	1	3%
#4 Instructor Asks Question:	5	17%
#5 Instructor Lectures:	11	37%
#6 Instructor Gives Directions:	0	0%
#7 Learner Response Specific:	2	7%
#8 Learner Initiates:	2	7%
#9 Learner Asks Question:	3	10%
#0 Silence or Confusion:	5	17%

Table 3-7.--Sample Data Matrix

The Instructional Verbal Analysis System
Copyright 1989 by Todd Hoover
Extended Printing of Data

Instructor: Donald Carter File Name
The matrix follows (R by C).....

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C0	ROW	TOTAL
R1	1	3	0	1	1	1	0	0	2	1	10	
R2	5	1	1	0	6	2	0	0	1	0	16	
R3	0	2	0	1	6	2	0	0	0	0	11	
R4	0	6	6	13	4	4	7	1	0	0	41	
R5	4	3	2	18	59	6	3	3	1	0	99	
R6	0	0	2	5	12	1	3	0	0	0	23	
R7	0	1	0	1	6	6	2	5	3	1	25	
R8	0	0	0	2	3	1	3	11	1	1	22	
R9	0	0	0	0	2	0	6	2	31	4	45	
R0	0	0	0	0	0	0	1	0	5	1	7	

Total number of entries = 299

The matrix allows for different questions to be asked i.e., how many times did students ask questions following directions given by the instructor.

Once the computer has created the matrix, the ratios are calculated; each ratio can have a value of from 0 to 100. The following are the calculations for each of the four ratios.

$$\text{Responsive Ratio} = \frac{1+2+3}{1+2+3+4+5+6}$$

$$\text{Dominant Ratio} = \frac{4+5+6}{1+2+3+4+5+6}$$

$$\text{Questioning Ratio} = \frac{4}{4+5}$$

$$\text{Initiative Ratio} = \frac{8+9}{7+8+9}$$

The Responsive and the Dominant ratios are inversely related to one another, as one increases the other decreases. The Responsive Ratio measures the percentage of time the instructor is responding to the student. The Dominant Ratio indicates the amount of time the instructor is controlling the pace of the class by lecturing, asking questions, or giving directions. The Questioning Ratio indicates the proportion of time the instructor spends lecturing versus the proportion of time the instructor spends questioning the students. Finally, the Initiative Ratio reflects the difference in student-initiated

communication compared to student response to instructor questioning.

The second screen of factors is printed (Figure 3-8) in sets and the percentages of each possible score are calculated within each category (Figure 3-9). Every four minutes IVA switches screens to allow the evaluator to rate the effective teaching factors. During the half hour evaluation session seven switches are performed.

Even though IVA has been used primarily as a self-assessment tool it should be useful in predicting teaching success as measured by the students and the administrative

Table 3-8.--Sample Factor Output

The Instructional Verbal Analysis System
Copyright 1990 by Todd Hoover
Extended Printing of Data

Instructor: File Name is:
The FACTOR data follows....

	Enthus.	Clarity	Orient.	Rapport	Organization
Set # 1	2	2	3	2	2
Set # 2	2	3	3	2	3
Set # 3	2	3	3	2	3
Set # 4	2	3	3	2	3
Set # 5	2	3	3	2	3
Set # 6	2	3	3	2	2
Set # 7	3	3	3	3	4

evaluator. Guskey[59] stated that successful instructors have four categories of characteristics in common. They plan and organize their lessons, have a positive regard for their

[59]Guskey, Improving Student Learning, 15.

Table 3-9.--Sample Factor Percentage Output

The Instructional Verbal Analysis System
Copyright 1990 by Todd Hoover
Extended Printing of Data

FACTOR: Enthusiasm

Percent of 1's chosen:	0%
Percent of 2's chosen:	100%
Percent of 3's chosen:	0%
Percent of 4's chosen:	0%
Percent of 5's chosen:	0%
Percent of 0's chosen:	0%

students, have a high degree of student participation during a class, and provide plenty of positive feedback. He further stated that these commonalities reflect teaching behaviors and practices, not personal characteristics. If pre-class planning is controlled, IVA will measure class organization, student participation, personal regard, and feedback.

Student Feedback Form

The Student Feedback Form (evaluation) was administered during the 9th week of each term. The evaluation (Appendix A) was divided into three parts: evaluation of teaching, evaluation of the course and evaluation of the facilities. Additionally, a global rating question asked the student to rate the instructor's overall performance at the beginning of the feedback form and two questions at the end of the form asked for overall performance ratings of the school and the course.

The teaching factors appearing on the student evaluation form were anchored in the factor analytical research related to components of effective instruction. Numerous studies have been conducted using factor analysis to identify characteristics of effective teaching. For example, Fenker, identified six factors that describe effective teaching.[60] He found the following six factors:

Factor 1: A good teacher factor. The best teachers are enthusiastic, intellectually stimulating, well prepared for class, coherent in presenting material, and aware of whether the class was following their discussion.

Factor 2: An evaluation factor. Items related to examinations have high correlations with this factor.

Factor 3: A factor related to course organization and items with emphasis on mechanical details.

Factor 4: An analytic/synthetic factor.

Factor 5: A factor related to instructor/individual student interactions.

Factor 6: A junk factor based on student classification items included in the questionnaire.

These six factors and those identified by Marsh and Guskey were used as a guide in developing the evaluation questionnaire. Twelve questions were crafted to measure teacher performance from the following categories: organization, enthusiasm, student understanding, rapport,

---

[60]R. M. Fenker, "The Evaluation of University Faculty and Administrators: A Case Study," Journal of Higher Education, XLVI:6 (November/December 1975)

and feedback. A sixth category was added--practical and relevant examples--that reflected the practitioner orientation of the school.

The pilot evaluation form was then submitted to the Assessment Committee, consisting of Associate Academic Deans and qualified central staff who rated the items as indicants of good or effective teaching. In addition, an outside evaluation was performed by a measurement specialist. Using Cronbach's alpha the student feedback form received a reliability score of .9083.

Independent Variables

The independent variables used for this study were the ratios produced using IVA. The Response Ratio corresponds to the categories that the faculty initiated such as response to a question or comment, or provision of feedback to a student. The Dominant Ratio reflected instructor directed communication. It should be noted that there was an inverse relationship between these two ratios. For example, if the Dominant Ratio were 81%, the Response ratio would be 19%. This reflected a situation in which the instructor dominated the communication taking place during the class session with little emphasis given to feedback.

The Questioning Ratio was a partial ratio, derived from the Dominant Ratio. This ratio reflected the proportion of time the instructor probed for understanding during a lecture. The greater the value of this ratio, the greater

the amount of time spent questioning to determine student understanding.

The Initiative Ratio measured the percentage of responses to an instructor initiated question versus responses initiated by the learner. This ratio was considered to be useful in gauging the degree to which the learners are active in participating in their own instruction. The ten individual categories were also used as independent variables. The additional factors of enthusiasm, clarity, task orientation, rapport, and organization were also used.

Dependent Variable

The dependent variable for this study was the global rating scores for faculty performance gathered from the student feedback form and the faculty observation evaluation form. The student global rating score represented the students' attitude or perception of the instructor's overall performance during the term.

Design

As noted earlier, the purpose of this investigation was to determine the effectiveness of IVA as a predictor of instructor success as rated by students on the global rating question of the student feedback form. The data set was analyzed using multiple regression in order to test the effects and the magnitudes of the independent variables on

the dependent variable. A significance level of .05 was used as the basis for rejecting the null hypotheses. The following null hypotheses were tested:

Null Hypothesis 1: There is no relationship between the IVA ratio scores and the student ratings.

Null Hypothesis 2: There is no relationship between the Dominant Ratio and the student ratings.

a. Instances of asking questions have no influence on student ratings.

b. Instances of lecturing or giving instructions have no influence on student ratings.

c. Instances of giving directions or organizing have no influence on student ratings.

Null Hypothesis 3: The Responsive Ratio has no influence on student ratings.

a. Instances of clarifying or answering questions have no influence on student ratings.

b. Instances of praise or encouragement have no influence on student ratings.

c. Instances of accepting and using ideas of the learner have no influence on student ratings.

Null Hypothesis 4: There is no relationship between the Questioning Ratio and student ratings.

Null Hypothesis 5: There is no relationship between the Initiative Ratio and student ratings.

a. Instances of learners initiating their own comments or responses have no influence on student ratings.

b. Instances of learners asking questions have no influence on student ratings.

Null Hypothesis 6: There is no relationship between Enthusiasm expressed by the faculty and student ratings.

Null Hypothesis 7: There is no relationship between Clarity of instruction by the faculty and student ratings.

Null Hypothesis 8: There is no relationship between Task Orientation of the faculty and student ratings.

Null Hypothesis 9: There is no relationship between Rapport with the students by faculty and student ratings.

Null Hypothesis 10: There is no relationship between Class organization by the faculty has and student ratings.

Chapter 4

RESULTS

The purpose of this study was to ascertain if the Instructional Verbal Analysis (IVA) software can be used to predict teaching performance as rated by students. IVA was designed to measure the verbal interactions between an instructor and his/her students. It was theorized that observations of certain interactions in classes could be used to predict teaching success as measured by satisfaction with the course as reported by the students.

During this study, the subjects were observed for 30 minutes, during which time verbal activity was recorded every three seconds using the IVA software and a computer. The ten categories used, over the 30 minutes, provided a time sampling of the verbal activity taking place. Every four minutes an additional group of categories was presented to the observer for consideration. This additional group of descriptors represented additional factors related to effective instruction. These effective instruction factors were selected based on the research discussed in Chapter 2.

Twenty-four faculty were observed during this study. Table 4-1 lists the means and standard deviations for all of the categories used by IVA, plus the mean student

rating of the faculty at the end of the course. The first four variables are means of ratios. The "Responsive" and "Dominant" ratios have an inverse relationship, that is, as one increases the other decreases. These ratios reflect the amount of time the instructor either responded to or directed the student. The "Questioning Ratio" and the "Initiative Ratio" measure the students involvement during the class session. The categories "Answers Questions" through "Silence" are reflected as percentages.

TABLE 4-1

Means and Standard Deviations of Variables

Label	Mean	Std Dev
Responsive Ratio	13.46	9.05
Dominant Ratio	85.54	9.05
Questioning Ratio	11.13	6.84
Initiative Ratio	42.88	28.26
Answers Questions	11.00	7.22
Praises	.04	.20
Uses Ideas	.08	.28
Asks Questions	13.54	17.97
Lectures	62.58	13.07
Gives Directions	.42	.83
Learner Response	9.50	6.16
Learner Initiates	2.88	5.55
Learner Questions	4.58	3.05
Silence	.92	2.65
Enthusiasm	2.37	.80
Clarity	2.15	.71
Task	2.70	.80
Rapport	1.88	.79
Organization	1.96	1.03
Rating	3.23	.45

The categories "Enthusiasm" through "Organization" were measured on a one to five scale, with one being equivalent

to high and five being equivalent to low. "Rating" was measured on a one to four scale with four representing a rating of excellent and one representing a rating of poor. The overall mean rating given to the faculty in this study by the students was 3.23.

Ten null hypotheses were tested to determine the effects and magnitudes of the independent variables on the dependent variable. The correlation matrix indicated high correlations, greater than plus or minus .5, among the variables Enthusiasm, Clarity, Task, Rapport, Organization and the dependent variable Rating. Plots of these independent variables and the dependent variable allow for a visual inspection of the strength of each of the relationships. Inspection of the plots reveal a strong linear relationship for the independent variables Enthusiasm, Clarity, Rapport and Organization (Figures 4-1 through 4-4).

Testing the Null Hypotheses

To test the null hypotheses regression analysis was used to determine the effects and magnitudes of the independent variables on the dependent variable. Ten hypotheses were tested.

Null Hypothesis 1: There is no relationship between the IVA ratio scores and the student ratings. Regression analysis was performed using the stepwise method on all of the IVA variables.

TABLE 4-2

CORRELATION MATRIX

	RESPON	DOMIN	QUEST	INIT	ANSWER	PRAISE	USESI	ASKSQ	LECTUR	DIRECT	LANSW	INITIA	LASKS	SILENC	ENTHUS	CLARITY	TASK	RAPPORT	ORGAN	RATING
RESPON	1.000																			
DOMIN	-1.000	1.000																		
QUEST	.078	-.078	1.000																	
INIT	.281	-.281	-.671	1.000																
ANSWER	.969	-.969	.040	.237	1.000															
PRAISE	-.011	.011	.027	-.059	-.029	-1.000														
USESI	-.339	.339	-.186	-.020	-.362	-.063	1.000													
ASKSQ	-.160	.160	-.146	.144	-.149	-.066	-.121	1.000												
LECTUR	-.705	.705	-.631	.079	-.598	-.075	.446	.107	1.000											
DIRECT	-.322	.322	.289	-.124	-.348	.150	.031	.255	-.136	1.000										
LANSW	.042	-.042	.904	-.753	.033	.155	-.200	-.144	-.557	.204	1.000									
INITIA	.269	-.269	-.186	.530	.077	-.034	-.132	-.013	-.357	.021	-.292	1.000								
LASKS	.639	-.639	-.137	.664	.519	.099	-.311	-.114	-.550	-.135	-.243	.600	1.000							
SILENC	.183	.183	-.174	.091	-.193	.007	-.106	.543	.067	.412	-.194	-.110	.071	1.000						
ENTHUS	-.357	.357	-.170	-.157	-.315	-.176	-.252	-.112	.307	.037	-.140	.055	-.266	-.049	1.000					
CLARITY	-.546	.546	-.214	-.088	-.478	-.044	-.280	.101	.451	.049	-.119	-.078	-.314	.043	.768	1.000				
TASK	-.459	.459	-.336	.018	-.464	-.149	-.049	.206	.427	-.186	-.245	.188	-.241	.000	.669	.626	1.000			
RAPPORT	-.364	.364	-.284	.100	-.266	.032	-.318	.180	.371	.022	-.218	-.135	-.169	.150	.647	.753	.501	1.000		
ORGAN	-.487	.487	-.201	-.049	-.396	-.198	-.180	.122	.480	.058	-.155	-.184	-.379	-.004	.651	.798	.367	.639	1.000	
RATING	.392	-.392	.231	-.120	.334	.105	.320	-.023	-.323	-.147	.199	-.054	.170	-.006	-.805	-.793	-.512	-.781	-.774	1.000

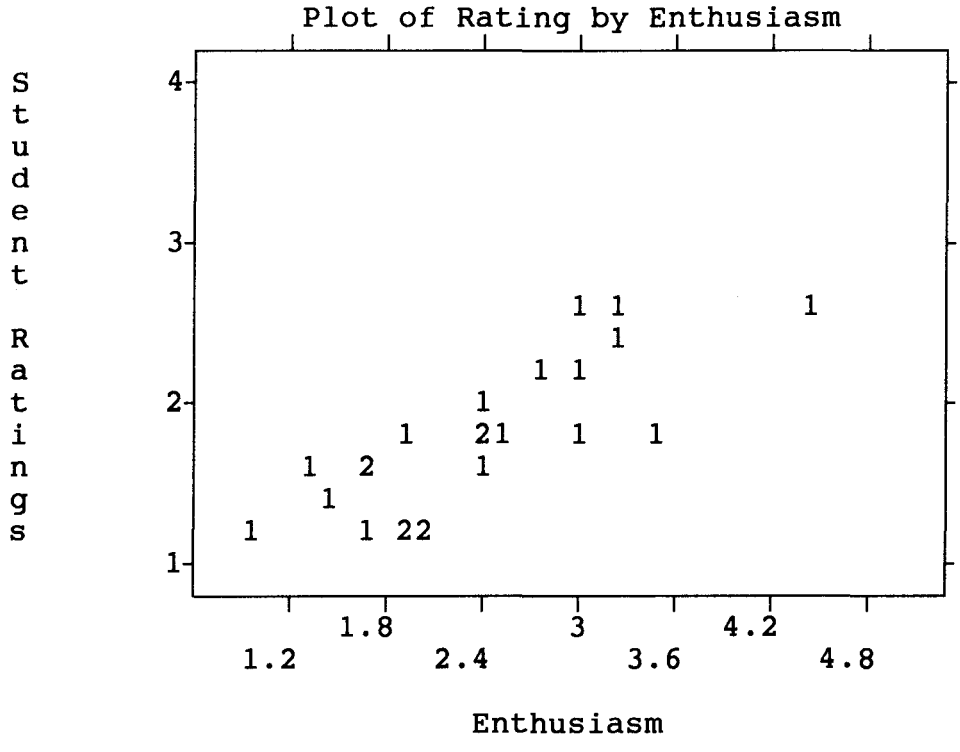

Figure 3. Plot of Rating by Enthusiasm

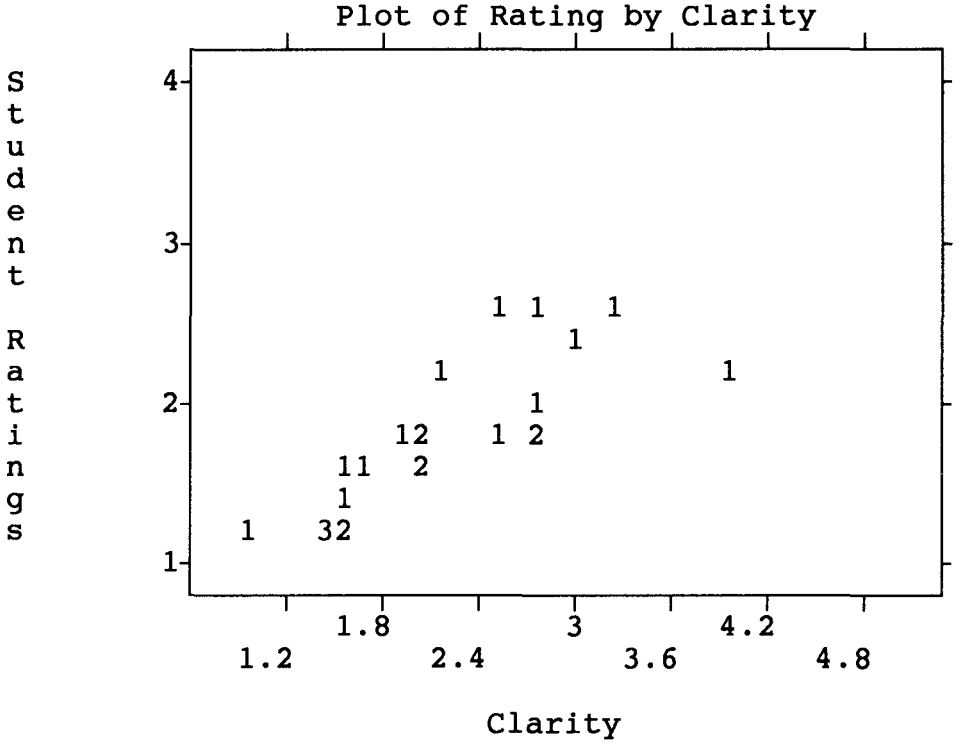

Figure 4. Plot of Rating by Clarity

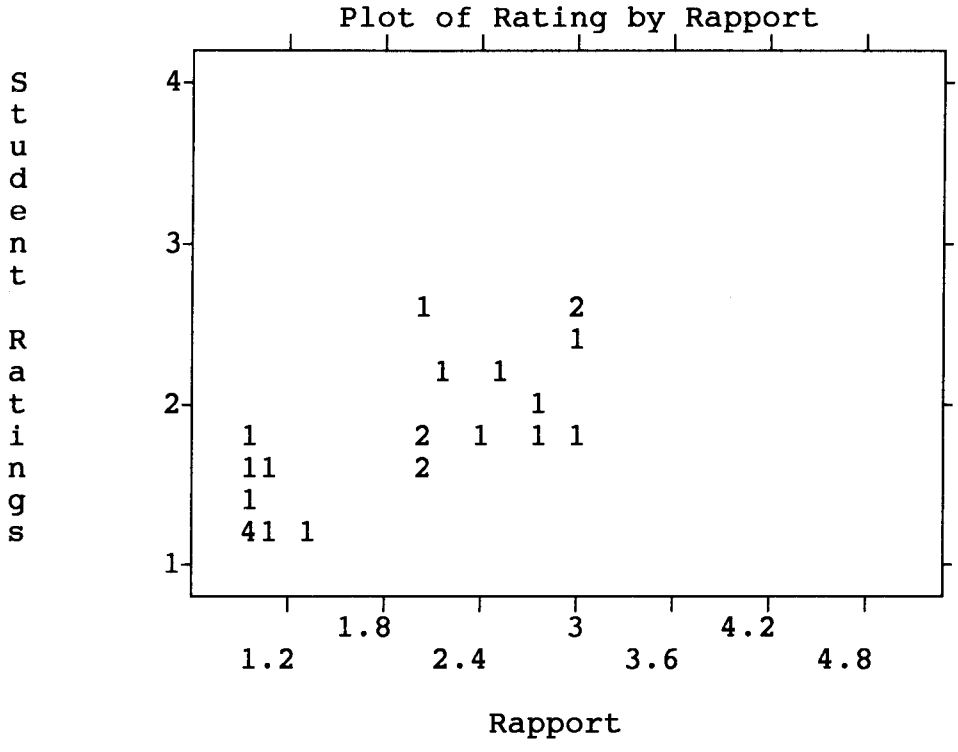

Figure 5. Plot of Rating by Rapport

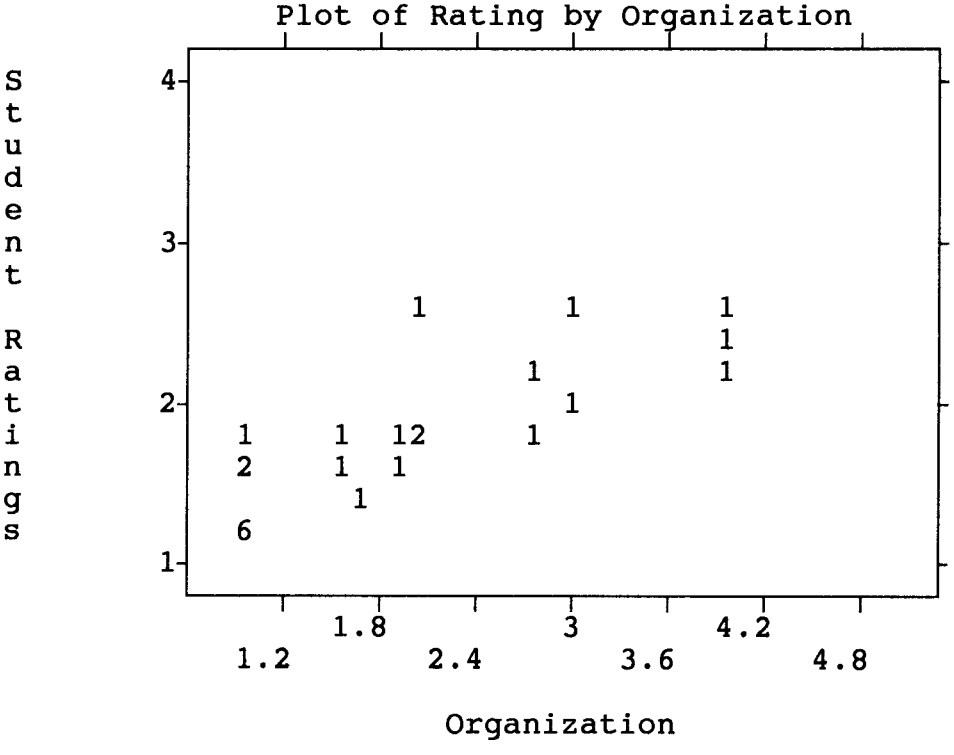

Figure 6. Plot of Rating by Organization

TABLE 4-3

REGRESSION STATISTICS FOR HYPOTHESIS ONE

Variable(s) Entered on Steps 1, 2 and 3

Enthusiasm, Rapport and Organization

Multiple R	.90084
R Square	.81151
Adjusted R Square	.78324
Standard Error	.20779

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	3	3.71785	1.23928
Residual	20	.86353	.04318

F = 28.70257 Signif F = .0000

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
ENTHUSIASM	-.216699	.077310	-.390293	-2.803	.0110
RAPPORT	-.188855	.078114	-.332226	-2.418	.0253
ORGANIZATION	-.132996	.059736	-.307297	-2.226	.0376
(Constant)	4.360473	.138101		31.575	.0000

The variables Enthusiasm, Rapport and Organization were found to have significant influence on the dependent variable Rating. The R Square value (.81) indicates the proportion of variance of the dependent variable accounted for by the independent variables. Thus, about 81% of the variance in student ratings is accounted for by Enthusiasm, Rapport, and Organization. The Adjusted R Square of .783 indicates that 78% of the variance in student ratings is accounted for by the three independent variables. The Adjusted R Square takes into account that it assumed that

the regression model always fits the data on which it was developed better than it will fit the population.[61] Therefore the Adjusted R Square is used to represent the population variables.

The results from the regression analysis produced the following equation: Student ratings = 4.360473 + (-.216699) (Enthusiasm) + (-.188855) (Rapport) + (-.132996) (Organization). This equation predicts the student ratings that faculty will receive given their scores on the three IVA variables Enthusiasm, Rapport, and Organization. An instructor who receives a score of one (a rating of high) on all three variables would receive a rating of 3.82 according to the prediction equation. Note that the beta weights for each of the variables is negative which is again, due to reverse scoring.

The casewise plot of standardized residuals (Figure 4-6) shows the residuals for each of the cases used. There are two cases that have residual values greater than +2 or -2. If there is a completely normal distribution with a mean of 0 and a standard deviation of 1, 95% of the cases will fall within the +2 or -2 range.[62] The cases with a value of greater than -2 were examined to determine if there were any errors in recording the data. No errors were discovered and

---

[61]M. J. Norusis, The SPSS Guide to Data Analysis (Chicago, IL: SPSS Inc., 1986), 346.

[62]Ibid., 353.

the cases were not eliminated from the data set. "As a general rule, outliers should be rejected out of hand only if they can be traced to causes such as error in recording the observations or in setting up the apparatus."[63]

Casewise Plot of Standardized Residuals

Case #	-3.0 O:.....:.....:O 3.0	RATING	*PRED	*ZRESID
1	. * . .	3.1	3.3977	-1.3844
2	. * .	3.5	3.4792	-.1406
3	. .* .	3.2	3.1239	.3664
4	. . * .	2.5	2.3019	.8569
5	. .* .	3.7	3.6356	.3101
6	. . * .	3.8	3.6356	.6951
7	. .* .	3.4	3.4034	.1761
8	. . * .	3.3	2.9583	1.4518
9	. * . .	2.5	2.9337	-2.3277
10	. . * .	2.7	2.5490	.6786
11	. . * .	2.9	2.7742	.5090
12	. .* .	3.7	3.6343	.2680
13	. * .	3.1	3.1090	.1493
14	. * . .	2.3	2.8150	-2.3821
15	. * .	3.0	2.9231	.1294
16	. .* .	3.2	3.1040	.3177
17	. *. .	3.4	3.5120	-.3467
18	. * . .	3.5	3.6183	-.6654
19	. *. .	3.8	3.8219	-.3461
20	. .* .	3.7	3.6133	.4654
21	. .* .	3.7	3.6052	.4561
22	. . * .	3.7	3.5788	.5834
23	. . * .	3.3	3.0852	.7931
24	. * . .	2.8	2.9274	-.6133
Case #	-3.0 O:.....:.....:O 3.0			

Figure 8. Casewise Plot of Standardized Residual

Based on the findings of the regression analysis the first null hypothesis is rejected.

---

[63]N. R. Draper and H. Smith, Applied Regression Analysis (New York, NY: John Wiley & Sons, Inc.), 153.

Null Hypothesis 2: There is no relationship between the Dominant Ratio and student ratings. This ratio consists of instructor led activities; asking questions, lecturing and giving directions.

Table 4-5 shows the regression analysis for the variables Dominant Ratio and student ratings. The resulting R Square was .154 indicating a weak linear relationship between the two variables. The significance of the F test was .0578 which exceeds the alpha of .05 and was not significant, consequently, we failed to reject the second null hypotheses.

TABLE 4-5

REGRESSION STATISTICS FOR HYPOTHESIS TWO

Variable(s) Entered on Step Number
1.. DOMINANT RATIO

Multiple R	.39246
R Square	.15402
Adjusted R Square	.11557
Standard Error	.41973

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	.70564	.70564
Residual	22	3.87575	.17617

F = 4.00543 Signif F = .0578

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
DOMIN	-.019353	.009670	-.392457	-2.001	.0578
(Constant)	4.886348	.831621		5.876	.0000

Null Hypothesis 2a: Instances of asking questions have no influence on student ratings. This variable represents the time spent by the instructor asking questions of the students during the class session.

Table 4-6 presents the regression analysis of the two variables Asking Questions and student ratings. The resulting R Square was .00051 indicating no linear relationship. The significance of the F test was .9167 that exceeds the alpha of .05 and was not significant, consequently, we failed to reject the null hypothesis 2a.

TABLE 4-6

REGRESSION STATISTICS FOR HYPOTHESIS TWO-A

Variable(s) Entered on Step Number
1.. ASKSQ_

Multiple R	.02255
R Square	.00051
Adjusted R Square	-.04492
Standard Error	.45622

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	.00233	.00233
Residual	22	4.57905	.20814

F = .01119 Signif F = .9167

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
ASKSQ_	-5.60008E-04	.005293	-.022552	-.106	.9167
(Constant)	3.238417	.117514		27.558	.0000

Null Hypothesis 2b: Instances of lecturing or giving instructions have no influence on student ratings. The lecturing variable represents the amount of time the instructor spends presenting information to the students.

Table 4-7 presents the regression analysis of the two variables Lecturing and student ratings. The resulting R Square was .10459 indicating an extremely weak linear relationship. The significance of the F test was .1232 which exceeds the alpha of .05 and is not significant, consequently, we failed to reject the null hypothesis 2b.

TABLE 4-7

REGRESSION STATISTICS FOR HYPOTHESIS TWO-B

Variable(s) Entered on Step Number

1.. LECTURE

Multiple R	.32340
R Square	.10459
Adjusted R Square	.06389
Standard Error	.43182

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	.47916	.47916
Residual	22	4.10222	.18646

F = 2.56971 Signif F = .1232

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
LECTURE	-.011048	.006892	-.323401	-1.603	.1232
(Constant)	3.922239	.440226		8.910	.0000

Null Hypothesis 2c: Instances of giving directions or organizing have no influence on student ratings. Table 4-8 presents the regression analysis of the two variables Giving Directions and student ratings. The resulting R Square was .023 indicating no linear relationship. The significance of the F test was .4943 which exceeds the alpha of .05 and was not significant, consequently, we failed to reject the null hypothesis 2c.

TABLE 4-8

REGRESSION STATISTICS FOR HYPOTHESIS TWO-B

Variable(s) Entered on Step Number

1.. GIVING DIRECTIONS

Multiple R	.14657
R Square	.02148
Adjusted R Square	-.02300
Standard Error	.45141

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	.09842	.09842
Residual	22	4.48296	.20377

F = .48300 Signif F = .4943

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
DIRECTIONS	-.078842	.113445	-.146570	-.695	.4943
(Constant)	3.263684	.103561		31.515	.0000

Null Hypothesis 3: There is no relationship between Responsive Ratio and student ratings.

Table 4-9 presents the regression analysis of the two variables Responsive Ratio and student ratings. The resulting R Square was .154 indicating almost no linear relationship. The significance of the F test was .0578 which exceeds the alpha of .05 and was not significant, consequently, we failed to reject the null hypothesis 3.

TABLE 4-9

REGRESSION STATISTICS FOR HYPOTHESIS THREE

Variable(s) Entered on Step Number

1.. RESPONSIVE RATIO

Multiple R	.39246
R Square	.15402
Adjusted R Square	.11557
Standard Error	.41973

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	.70564	.70564
Residual	22	3.87575	.17617

F = 4.00543 Signif F = .0578

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
RESPON	.019353	.009670	.392457	2.001	.0578
(Constant	2.970370	.155813		19.064	.0000

Null Hypothesis 3a: Instances of clarifying or answering questions have no influence on student ratings.

Table 4-10 presents the regression analysis of the two variables Answering Questions and student ratings. The resulting R Square was .11178 indicating almost no linear relationship. The significance of the F test was .1103 which exceeds the alpha of .05 and was not significant, consequently, we failed to reject the null hypothesis 3A

TABLE 4-10

REGRESSION STATISTICS FOR HYPOTHESIS THREE-A

Variable(s) Entered on Step Number

1.. ANSWERING QUESTIONS

Multiple R	.33434
R Square	.11178
Adjusted R Square	.07141
Standard Error	.43008

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	.51212	.51212
Residual	22	4.06926	.18497

F = 2.76872 Signif F = .1103

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
ANSWERING	.020658	.012415	.334339	1.664	.1103
(Constant)	3.003592	.162351		18.501	.0000

Null Hypothesis 3b: Instances of praise or encouragement have no influence on student ratings.

Table 4-11 presents the regression analysis of the two variables Praise and student ratings. The resulting R Square was .01094 indicating no linear relationship. The significance of the F test was .6267 which exceeds the alpha of .05 and was not significant, consequently, we failed to reject the null hypothesis 3b.

TABLE 4-11

REGRESSION STATISTICS FOR HYPOTHESIS THREE-B

Variable(s) Entered on Step Number

1.. PRAISE

Multiple R	.10460
R Square	.01094
Adjusted R Square	-.03402
Standard Error	.45384

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	.05012	.05012
Residual	22	4.53126	.20597

F = .24335 Signif F = .6267

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
PRAISE	.228696	.463596	.104597	.493	.6267
(Constant)	3.221304	.094631		34.041	.0000

Null Hypothesis 3c: Instances of accepting and using ideas of the learner have no influence on student ratings.

Table 4-11 presents the regression analysis of the two variables Using Ideas and student ratings. The resulting R Square was .10261 indicating no linear relationship. The significance of the F test was .1270 which exceeds the alpha of .05 and was not significant, consequently, we failed to reject the null hypotheses 3c.

TABLE 4-11

REGRESSION STATISTICS FOR HYPOTHESIS THREE-C

Variable(s) Entered on Step Number

1.. USING IDEAS

Multiple R	.32032
R Square	.10261
Adjusted R Square	.06181
Standard Error	.43229

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	.47007	.47007
Residual	22	4.11131	.18688

F = 2.51541 Signif F = .1270

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
USING IDEAS	.506364	.319270	.320321	1.586	.1270
(Constant)	3.188636	.092165		34.597	.0000

Null Hypothesis 4: There is no relationship between the Questioning Ratio and student ratings.

Table 4-12 presents the regression analysis of the two variables Questioning Ratio and student ratings. The resulting R Square was .05326 indicating no linear relationship. The significance of the F test was .2780 which exceeds the alpha of .05 and was not significant, consequently, we failed to reject the null hypothesis 4.

TABLE 4-12

REGRESSION STATISTICS FOR HYPOTHESIS FOUR

Variable(s) Entered on Step Number

1.. QUESTIONING RATIO

Multiple R	.23077
R Square	.05326
Adjusted R Square	.01022
Standard Error	.44402

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	.24399	.24399
Residual	22	4.33740	.19715

F = 1.23755 Signif F = .2780

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
QUESTIONING	.015054	.013532	.230773	1.112	.2780
(Constant)	3.063358	.175724		17.433	.0000

Null Hypothesis 5: There is no relationship between the Initiative Ratio and student ratings.

Table 4-13 presents the regression analysis of the two variables Initiative Ratio and student ratings. The resulting R Square was .01442 indicating no linear relationship. The significance of the F test was .5762 which exceeds the alpha of .05 and was not significant, consequently, we failed to reject the null hypothesis 5.

TABLE 4-13

REGRESSION STATISTICS FOR HYPOTHESIS FIVE

Variable(s) Entered on Step Number

1.. INITIATIVE RATIO

Multiple R	.12008
R Square	.01442
Adjusted R Square	-.03038
Standard Error	.45304

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	.06606	.06606
Residual	22	4.51533	.20524

F = .32184 Signif F = .5762

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
INITIATIVE	-.001897	.003343	-.120076	-.567	.5762
(Constant)	3.312152	.170582		19.417	.0000

Null Hypothesis 5a: Instances of learners initiating their own comments or responses have no influence on student ratings.

Table 4-14 presents the regression analysis of the two variables Learners Initiating and student ratings. The resulting R Square was.00290 indicating no linear relationship. The significance of the F test was .8027 which exceeds the alpha of .05 and was not significant, consequently, we failed to reject the null hypothesis 5a.

TABLE 4-14

REGRESSION STATISTICS FOR HYPOTHESIS FIVE-A

Variable(s) Entered on Step Number

1.. LEARNER INITIATING

Multiple R	.05384
R Square	.00290
Adjusted R Square	-.04242
Standard Error	.45568

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	.01328	.01328
Residual	22	4.56810	.20764

F = .06395 Signif F = .8027

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
INITIA	-.004329	.017118	-.053837	-.253	.8027
(Constant)	3.243279	.105232		30.820	.0000

Null Hypothesis 5b: Instances of learners asking questions have no influence on student ratings.

Table 4-15 presents the regression analysis of the two variables Learners Asking Question and student ratings. The resulting R Square was .02876 indicating no linear relationship. The significance of the F test was .4282 which exceeds the alpha of .05 and was not significant, consequently, we failed to reject the null hypothesis 5b.

TABLE 4-15

REGRESSION STATISTICS FOR HYPOTHESIS FIVE-B

Variable(s) Entered on Step Number

1.. LEARNER ASKS QUESTION

Multiple R	.16960
R Square	.02876
Adjusted R Square	-.01538
Standard Error	.44973

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	.13178	.13178
Residual	22	4.44961	.20225

F = .65154 Signif F = .4282

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
LASKS	.024825	.030755	.169599	.807	.4282
(Constant)	3.117054	.168216		18.530	.0000

Null Hypothesis 6: There is no relationship between Enthusiasm expressed by the faculty and student ratings.

Table 4-16 presents the regression analysis of the two variables Enthusiasm and student ratings. The resulting R Square was .64844 indicating a linear relationship. The significance of the F test was .00000 which did not exceed the alpha of .05 and was significant, consequently, the null hypothesis was rejected.

TABLE 4-16

REGRESSION STATISTICS FOR HYPOTHESIS SIX

Variable(s) Entered on Step Number

1.. ENTHUSIASM

Multiple R	.80526
R Square	.64844
Adjusted R Square	.63246
Standard Error	.27058

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	2.97074	2.97074
Residual	22	1.61065	.07321

F = 40.57761 Signif F = .0000

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
ENTHUSIASM	-.447094	.070187	-.805256	-6.370	.0000
(Constant)	4.291563	.175439		24.462	.0000

Null Hypothesis 7: There is no relationship between Clarity of instruction by the faculty and student ratings.

Table 4-17 presents the regression analysis of the two variables Clarity and student ratings. The resulting R Square was .62866 indicating a linear relationship. The significance of the F test was .00000 which did not exceed the alpha of .05 and was significant, consequently, the null hypothesis was rejected.

TABLE 4-17

REGRESSION STATISTICS FOR HYPOTHESIS SEVEN

Variable(s) Entered on Step Number

1.. CLARITY

Multiple R	.79288
R Square	.62866
Adjusted R Square	.61178
Standard Error	.27808

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	2.88014	2.88014
Residual	22	1.70124	.07733

F = 37.24516 Signif F = .0000

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
CLARITY	-.497368	.081497	-.792882	-6.103	.0000
(Constant)	4.299139	.184023		23.362	.0000

Null Hypothesis 8: There is no relationship between Task orientation of the faculty and student ratings.

Table 4-18 presents the regression analysis of the two variables Task and student ratings. The resulting R Square was .26246 indicating a weak linear relationship. The significance of the F test was .0105 which did not exceed the alpha of .05 and was significant, consequently, the null hypothesis was rejected.

TABLE 4-18

REGRESSION STATISTICS FOR HYPOTHESIS EIGHT

Variable(s) Entered on Step Number

1.. TASK

Multiple R	.51231
R Square	.26246
Adjusted R Square	.22894
Standard Error	.39190

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	1.20245	1.20245
Residual	22	3.37893	.15359

F = 7.82907 Signif F = .0105

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
TASK	-.287000	.102571	-.512313	-2.798	.0105
(Constant)	4.004656	.287896		13.910	.0000

Null Hypothesis 9: There is no relationship between Rapport with the students by faculty and student ratings.

Table 4-19 presents the regression analysis of the two variables Rapport and student ratings. The resulting R Square was .61008 indicating a linear relationship. The significance of the F test was .00000 which did not exceed the alpha of .05 and was significant, consequently, the null hypothesis was rejected.

TABLE 4-19

REGRESSION STATISTICS FOR HYPOTHESIS NINE

Variable(s) Entered on Step Number

1.. RAPPORT

Multiple R	.78108
R Square	.61008
Adjusted R Square	.59236
Standard Error	.28495

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	2.79503	2.79503
Residual	22	1.78635	.08120

F = 34.42255 Signif F = .0000

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
RAPPORT	-.444008	.075678	-.781079	-5.867	.0000
(Constant)	4.065753	.153734		26.447	.0000

Null Hypothesis 10: There is no relationship between Class organization by the faculty and student ratings.

Table 4-20 presents the regression analysis of the two variables Organization and student ratings. The resulting R Square was .59849 indicating a linear relationship. The significance of the F test was .00000 which did not exceed the alpha of .05 and is significant, consequently, the null hypothesis was rejected.

TABLE 4-20

REGRESSION STATISTICS FOR Hypothesis NINE

Variable(s) Entered on Step Number

1.. ORGAN

Multiple R	.77362
R Square	.59849
Adjusted R Square	.58024
Standard Error	.28916

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	2.74192	2.74192
Residual	22	1.83947	.08361

F = 32.79326 Signif F = .0000

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
ORGANIZATION	-.334819	.058468	-.773622	-5.727	.0000
(Constant)	3.886380	.128796		30.175	.0000

Between Group Analysis

As part of its mission, the Graduate School of Management seeks from its adjunct faculty a performance rating of 3.4 or better in order to be classified as an excellent instructor. For a new faculty member to be retained, a first time rating of 3.0 or better is sought, with a three term goal of increasing his or her rating to 3.4 or better. It should be noted that faculty who receive a rating of less than 3.0 will be considered on a individual basis as to whether or not the individual will continue to teach in the program. Regression analysis was used to explore for relationships among the faculty. Using the rating requirements of the Graduate School the faculty were divided into groups based on their ratings. The following regression statistics (Table 4-21) are for the group of faculty who received ratings of 3.4 or greater.

Table 4-21.

Regression Statistics for Group with Rating of > 3.4

Variable(s) Entered on Step Number
1.. CLARITY

Multiple R	.75818
R Square	.57484
Adjusted R Square	.52760
Standard Error	.09423

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	.10805	.10805
Residual	9	.07991	.00888

F = 12.16843 Signif F = .0068

----------Variables in the Equation----------

Variable	B	SE B	Beta	T	Sig T
CLARITY	-.373848	.107171	-.758181	-3.488	.0068
(Constant)	4.209099	.170736		24.653	.0000

Variable(s) Entered on Step Number
2.. INITIATES

Multiple R	.87340
R Square	.76283
Adjusted R Square	.70353
Standard Error	.07465

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	2	.14338	.07169
Residual	8	.04458	.00557

F = 12.86522 Signif F = .0032

----------Variables in the Equation----------

Variable	B	SE B	Beta	T	Sig T
CLARITY	-.362912	.085012	-.736003	-4.269	.0027
INITIA_	-.007579	.003010	-.434141	-2.518	.0359
(Constant)	4.218103	.135304		31.175	.0000

The two variables Clarity and Initiates account for approximately 76% of the variance in student ratings for faculty with a student rating of 3.4 or greater.

Clarity was defined as the ability of an instructor to deliver a lesson in a nonconfusing and concise manner. Examples are used to support and reinforce the concepts being presented. The variable Initiates refers to incidences of students initiating discussion. Neither of these variables appeared in the regression equation for the entire group when all of the IVA variables were entered.

Examination of the correlation matrix revealed a high correlation between Clarity and Enthusiasm (.768), Rapport (.753) and Organization (.798). There was no linear relationship found between Initiates and the other independent variables. The distinguishing factors among the faculty who received a rating of 3.4 or better were their differences in presenting information in a clear and concise manner and instances of students initiating discussion.

Regression analysis was performed on the data for those faculty who received a rating of less than 3.4 (see table 4-22).

Table 4-22.

Regression Statistics for Group with Rating of < 3.4

Variable(s) Entered on Step Number
1.. ENTHUS

Multiple R	.56419
R Square	.31831
Adjusted R Square	.25634
Standard Error	.28263

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	1	.41030	.41030
Residual	11	.87870	.07988

F = 5.13637 Signif F = .0446

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
ENTHUS	-.288782	.127421	-.564190	-2.266	.0446
(Constant)	3.742800	.380046		9.848	.0000

For those instructors with student ratings of less than 3.4, Enthusiasm was found to be the most important factor in determining their rating. The magnitude of Enthusiasm was much less for this group with it accounting for 31% of the variance in student ratings.

For faculty who received a rating of less than or equal to 2.8 the results (Table 4-23) were as follows.

Table 4-23.

Regression Statistics for Group with Rating of < = 2.8

Variable(s) Entered on Step Number
1.. LEARNER ANSWERS

Multiple R	.86503
R Square	.74828
Adjusted R Square	.68534
Standard Error	.12334

Analysis of Variance

	DF	Sum of Squares	Mean Square
regression	1	.18088	.18088
Residual	4	.06085	.01521

F = 11.98040 Signif F = .0261

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
LANSW	.034630	.010043	.856029	3.448	.0261
(Constant)	2.262807	.110850		20.413	.0000

Faculty who fell in this bracket had a decreased amount of student activity in their classes. For this group there was a high correlation between the lack of instances of the learner answering questions and ratings. The R Square was

.74828 indicating that approximately 74% of the instructors rating was accounted for by the variable.

Summary of Results

Regression analysis was used to determine the effects and magnitudes of the independent variables on the dependent variable. Ten null hypotheses were tested to determine the relationships of the IVA variables to the student ratings.

Three variables Enthusiasm, Rapport, and Organization were found to have a significant influence on the dependent variable Student Ratings. The analysis yields an F test vallue of significance at the .0014 level. Examination of the prediction equation reveald that an instructor would receive a student rating of 3.82 if he or she received a high rating (one) on each of the three variables.

When an analysis was done on each of the individual variables of IVA, only Enthusiasm, Clarity, Task, Rapport and Organization were found to be significant. A summary of the regression analysis for each of the individual hypotheses was presented in table 4-24.

Additional analysis were performed in an effort to determine if any differences existed between faculty groups based on student ratings. It was determined that the most influential variables for faculty who received a rating of 3.4 or better was Clarity and instances of Initiating Discussion. For faculty who received a rating less than 3.4, Enthusiasm seemed to be the most influential variable.

Finally, for faculty who received a rating of less than 2.8 there was a high correlation between the lack of instances of the learner answering questions and student ratings.

Table 4-24.

Results of the Regression Analysis

Null Hypotheses	Results
One--IVA	Rejected
Two--Dominant Ratio	Failed to reject
Two A--Asking Questions	Failed to reject
Two B--Lecturing	Failed to reject
Two C--Directions	Failed to reject
Three--Responsive Ratio	Failed to reject
Three A--Answering	Failed to reject
Three B--Praise	Failed to reject
Three C--Ideas	Failed to reject
Four--Questioning Ratio	Failed to reject
Five--Initiative Ratio	Failed to reject
Five A--Initiating	Failed to reject
Five B--Asking Questions	Failed to reject
Six--Enthusiasm	Rejected
Seven--Clarity	Rejected
Eight--Task	Rejected
Nine--Rapport	Rejected
Ten--Organization	Rejected

Chapter 5

DISCUSSION

Introduction

The purpose of this study was to determine the effectiveness of IVA as a predictive tool in measuring student satisfaction with the instructor. This chapter integrates the study's findings in an attempt to provide direction for one of education's purposes of providing excellence in teaching. The results of the investigation are discussed, conclusions are drawn, and needs for future research are examined.

Results

Twenty-four adjunct faculty members were evaluated using IVA prior to their students completing the student feedback form. For these instructors ten null hypotheses were tested using regression analysis to determine the effects and magnitudes of the independent variables on the dependent variable. The results related to each are as follows:

Hypothesis 1: There is no relationship between the IVA ratio scores and student ratings.

Findings

The variables Enthusiasm, Rapport and Organization demonstrated significant influence on the dependent variable Rating. The F ratio was 28.70 and significant at the .0000 level. About 81% of the variance in the student ratings is accounted for by these three variables.

Discussion

It is interesting to note that the only significant variables used by IVA are those from the second screen. IVA's second screen is composed of the five variables Enthusiasm, Rapport, Organization, Clarity and Task which were all derived from the literature as main categories of effective instruction. The other ten categories and four ratios, which were derived from the original work of Flanders, had no effect on the dependent variable.

One explanation for the lack of influence is the character of the first ten variables. These variables are descriptive in nature, they describe the verbal activity taking place in a classroom. Upon examination of the raw data, one can get a picture of this activity. Most is very teacher directed in the form of lecture or questioning. In the classrooms of extremely effective faculty, learner initiation of comments and questions is high. The questioning activity is inquisitive, students are interested in the topic and are exploring alternatives as opposed to asking questions of clarification. In situations where

clarification is needed, the instructor is very proactive in probing the students for understanding and uses questions to clarify. While all of this activity is important, it does not appear to be of value in the evaluation of student satisfaction with the instructor.

A second explanation deals with the nature of the data itself. When using IVA's screen one to gather data, a time sampling of the verbal activity is being taken. Collecting frequency data on verbal activity is different than measuring the effectiveness of the activity. For example, significant feedback to the student by the instructor is extremely ineffective if it is sarcastic in nature. Questioning students has little effect if the questions are inappropriate.

The second screen variables are rated using a one to five measure and being used by the evaluator to measure the effectiveness of the activity taking place in the classroom. Because of the difference in the type and measure of the IVA variables there may be some interference between the two. It may be that future versions of IVA must change the method of measurement of the original variables if it is to be used for predictive purposes.

Finally, some of the variables may be inappropriate for a graduate education setting. In courses that are highly content-based it is extremely rare for an instructor to use student's ideas and to and build upon those ideas. In

addition, the type of praise that an adult receives is different than the praise given to a primary grade student.

Hypothesis 2: There is no relationship between the Dominant Ratio and student ratings.

2a: Instances of asking questions have no influence on student ratings.

2b: Instances of lecturing or giving instructions have no influence on student ratings.

2c: Instances of giving directions or organizing have no influence on student ratings.

Findings

The null hypothesis for each of the above was not rejected. Each of the F ratios were not significant at the .05 level.

Discussion

Although the hypothesis of no difference was not rejected, the Dominant Ratio's F ratio was significant at the .06 level. Again, analysis of the raw data reveals a similar pattern with the successful faculty. Most of those who had received a rating of 3.3 or better had dominant ratios in the range of 75 to 85. Removal of one case with a high ratio would cause the variable to be significant. Even though the courses are content oriented the best instructors do not spend all of the class time lecturing without

significant student involvement. For those who have a higher dominant ratio, personality seems to play an important part in their success.

Hypothesis 3: There is no relationship between the Responsive Ratio and student ratings.

3a: Instances of clarifying or answering questions have no influence on student ratings.

3b: Instances of praise or encouragement have no influence on student ratings.

3c: Instances of accepting and using ideas of the learner have no influence on student ratings.

Findings

The null hypothesis for each of the above was not rejected. Each of the F ratios were not significant at the .05 level.

Discussion

The Responsive Ratio and the Dominant Ratio have an inverse relationship with one another. As one increases, the other decreases, and their total will always equal 100. As was discussed earlier, the faculty with ratings of 3.3 or better will have a responsive ratio of between 15 and 25. With the exception of one faculty member, all of the unsuccessful faculty members had responsive ratios of less than 10. While it was not statistically significant,

patterns in the data do show that for the most part extremes in either the Responsive or Dominant Ratios lead to unsuccessful performance.

Hypothesis 4: There is no relationship between the Questioning Ratio and student ratings.

Finding

The null hypothesis for the above was not rejected. The F ratios were not significant at the .05 level.

Hypothesis 5: There is no relationship between the Initiative Ratio and student ratings.

5a: Instances of learners initiating their own comments or responses have no influence on student ratings.

5b: Instances of learners asking questions have no influence on student ratings.

Findings

The null hypothesis for each of the above was not rejected. Each of the F ratios were not significant at the .05 level.

Discussion

Both hypothesis 4 and 5 are comprised of data from the other variables which were themselves not significant. As previously stated, the frequencies of occurrence of each of the components does not seem to be an indicator of performance. Changing the scales might lead to different

results in future research.

Hypothesis 6: There is no relationship between Enthusiasm expressed by the faculty and student ratings.

Hypothesis 7: There is no relationship between Clarity of instruction by the faculty and student ratings.

Hypothesis 8: There is no relationship between Task orientation of the faculty and student ratings.

Hypothesis 9: There is no relationship between Rapport with the students by faculty as no influence on student ratings.

Hypothesis 10: Class organization by the faculty and student ratings.

Findings

Each of the above independent variables has a significant F ratio at the .00 level, except Task which is significant at the .01 level. The effects of each of the following independent variables on the dependent variable is as follows: Enthusiasm--.65, Clarity--.63, Task--.26, Rapport--.61, and Organization--.60.

Discussion

With the exception of Task, it is interesting to note the degree of influence each has on the dependent variable student rating. When all of the variables were analyzed

Task and Clarity were dropped from the regression equation. Examination of the correlation matrix reveals a high degree of correlation between Enthusiasm and Clarity (R = .768). When analysis was performed on faculty who had received a rating of equal to or greater than 3.4, Clarity along with Initiates were the significant variables. This is the first occurrence of an original IVA variable either individually or in conjunction with another variable. Together they account for .76283 of the variance in student ratings.

Because of the correlation between Clarity and Enthusiasm it is with caution that any definitive statements are made concerning this analysis. What seems to be occurring is that superior faculty do a superior job in providing students with clear and concise information on the particular topic being discussed. Additionally, students of superior faculty initiate more discussion concerning the topic being addressed. For example, during the evaluation of the Quality Management instructor, students were witnessed initiating discussions comparing their own company policies with the quality theories and how organizational changes could be implemented within their own departments or companies. These student initiated discussions continue for greater duration than those in classes of less than superior faculty.

Between Group Analysis

Excellent faculty, within the MBA program, are by

definition those who receive a mean student rating of 3.4 or better. Faculty who fall between 3.0 and 3.39 are performing adequately but with room for improvement, and those below 3.0 need major improvements in their teaching skills. The goal for all faculty is to receive a student rating of 3.4 or better each term they teach. Very few instructors ever receive a rating of 4.0, yet, a 4.0 rating term after term is the ultimate goal that all should strive to attain.

Feedback from IVA can be instrumental in achieving higher ratings for all faculty. Those in the 3.4 or greater category have room for improvement in their clarity of presentation. When regressions was performed on this group, clarity and instances of students initiating discussion were the two significant variables. Increases in rating at this level are difficult but this insight provides instructors with the means of improving their performance.

For this research, clarity has been defined as the ability to present information in a clear and concise manner. Instructors need to ensure their presentations flow from point to point. They must not digress from the lecture topic and their discussion must be complete. The instructor must be able to "read" each student to insure there is complete understanding of the material.

Those instructors wishing to improve must not only provide clear and concise instruction, but they need to

provide opportunity for their students to become more involved in the direction of the class discussions. Students of excellent instructors seem to be excited about to course, they are interested in making a contribution. For example, during an observation in a Quality Management course the students were the driving force during the discussions taking place. The instructor would ask a question to provide a direction for the group. Students would each address the question from their own perspective, which in turn would lead to further discussion. The instructor would intercede only to become part of the discussion or to bring some closure. These students were taking an active part in their education.

Research has indicated that enthusiasm is an integral part of effective classroom performance. IVA has shown that for faculty who received a student rating of less than 3.4, enthusiasm is the main difference in performance. The solutions seems simple enough, to improve performance increase the level of genuine enthusiasm. The factors used to rate an instructor on enthusiasm are: speaking expressively, moving about the classroom, and the use of appropriate humor. Enthusiasm should be a natural expression of the enjoyment of teaching.

For instructor with mean student ratings of less than 3.0, a single distinguishing factor was detected during the analysis. The lack of students answering question accounted

for 75% of the variance among the faculty. This would seem to indicate that faculty were not asking enough or the appropriate questions during their lectures. This lack of student participation is a major contributor of poor student ratings.

Recommended Design Improvements

There are two major areas in which improvements can be made in the implementation of IVA: (1) in the obtrusiveness of the evaluator, and (2) the design of IVA, if it is to be used as an predictive tool or an evaluative instrument.

Obtrusiveness of Evaluator

Using a laptop computer provides a convenient way to capture information. During pilot testing, an attempt was made to capture data using paper and pencil. This method was clearly inferior to the use of a computer. Although the laptop was clearly superior to the pencil and paper method, it also created a set of its own unique problems. Although students are used to seeing computers, a certain amount of distraction does occur. When depressing the keys on some computers, a clicking sound is emitted that clearly distracts those close at hand. Disk drive noise can also cause distractions. In one case an instructor stopped class and asked if anybody heard that "strange noise." Closing the lid on some laptops while the power is on causes the speaker to beep continuously. There may even be a certain

amount of intimidation of the instructor when seeing an evaluator with a computer in class.

New technology will clearly solve both the problems stated above. Palmtop computers will allow evaluators to conveniently capture data while remaining as unobtrusive as possible. Until this technology is practical, notebook computers provide a good compromise. Although not as small as palmtops they are much less conspicuous.

Design of IVA

Screen one categories of IVA, which are based on Flanders research, require the evaluator to record the frequency of occurrence every three seconds. The results of this research show that this is not necessary if IVA is to be used as a predictor of instructor success. Screen two categories are measured every four minutes and use a Likert scale as opposed to frequencies. This leads this researcher to believe that screen one categories should be adjusted to provide the same evaluative measures used with screen two categories or possibly not used at all. With this change the evaluator would be assessing the quality of the lecture and the quality of the instructor's questions. This may provide additional information and provide better insights. In addition, the timing interval should be the same for both screens.

A second change would eliminate the two screens. The type of screens on laptops make it difficult for an

evaluator to notice the automatic switch even when a low tone is emitted as it occurs. The restructuring of IVA will accomplish this goal. Changing the format from frequencies to the Likert scale will allow the screens to be combined. Evaluating the degree of accomplishment of each category will allow the evaluator to concentrate on the effectiveness of the instructor instead of the frequency of his or her actions.

Another change that should occur is in the number of categories that equate with effective instruction. Presently, the five descriptors being used were derived from factor analysis of several more specific descriptors. The general nature of the factors could cause several different evaluators to apply their own biases to the meaning of each, even after extensive training. This can cause inter-rater reliability problems. One example of a program that is using a multitude of factors is the STAR[64] program used by the State of Louisiana.

STAR (System for Teaching and Learning Assessment and Review) was developed for the State of Louisiana in response to two laws: the Louisiana Teaching Internship Law and the Children First Act. These laws require on-the-job assessment of teaching for both beginning and existing teachers. The development of STAR was a cooperative venture

---

[64]C. D. Ellett, "STAR: System for Teaching and Learning Assessment and Review" 1990.

between the State, local school districts and colleges and universities.

STAR's developers consider it a second generation assessment program that measures more than just teacher skills. The common themes of STAR reflect its comprehensive nature. They are: (1) all children can learn; (2) student self responsibility for learning and learning to learn; (3) learner individual differences; (4) teaching and learning as a total process; (5) time; (6) quality learning environment; (7) physical environment; (8) thinking skills; (9) active involvement and engagement and; (10) knowledge of pedagogy, content and curriculum.

STAR consists of 140 assessment indicators that operationalize 23 teaching and learning components. Each of the indicators relate to components of effective instruction and student learning. Further analysis of STAR's 23 teaching and learning components reveal remarkable similarities with certain categories of IVA.

While STAR is an extremely comprehensive evaluation program, 140 assessment indicators does seem extreme. Expanding IVA's categories seems to utilize a more reasonable approach. The five descriptive categories should be expanded to approximately 20. These new indicators would be reflective of the broader descriptors presently used.

The following illustrates a possible new design:

Enthusiasm

Speaks expressively
Moves while lecturing
Uses humor

Rapport

Shows concern for students
Shows respect for students
Provides positive feedback

Clarity

Uses practical examples
Uses multiple examples
Stresses important points
Uses precise explanations

Task

Proceeds at moderate pace
Probe for student understanding
Digresses from topic
Involves students in presentation

Organization

Objectives stated clearly
Presents outline of lecture
Close topic integration
Explains how each topic fits
Smooth topic transitions

Each of the subindicators would be rated using the present one to five scale. The evaluator would rate the instructor every five minutes in all of the categories witnessed. A zero rating would be used if there Were "no rating" in any of the categories.

Finally, one additional area needs to be addressed by IVA, that of planning. For this research, faculty planning

was held constant with the entire process being very carefully documented prior to the beginning of the term. This is not usually the case in most instances. Programs such as STAR address this issue with evaluation taking place on small "chunks" of the curriculum.

The task of curriculum planning is one of the areas that education can learn from the quality concepts being put forth by business. Genichi Taguchi, a Japanese engineer and quality champion, states that achieving quality systems cannot be accomplished through inspection, or in this case evaluation, but systems must be designed so the product (in this case the curriculum) functions correctly from the beginning.[65] Curriculum design and lesson planning need to be considered the framework of the entire process. Instructors need to be evaluated during the entire process not just a few selected points in time.

Implications for Further Research

The results of this study suggest several research initiatives that should take place to resolve issues raised and to extend them as well.

1. The changes to IVA described above need to be implemented and a replication of this study needs to be performed. During this study, only one evaluator was used, eliminating the inter-rater reliability problems.

---

[65]B. Gunter, "A Perspective on the Taguchi Methods" Quality Progress (June 1987): 44.

The addition of several new factors along with extensive training should make it easier for several evaluators to make consistent evaluative decisions. This should allow for a larger sample size to be used, thus providing additional knowledge.

2. One issue this research did not resolve is the relationship between the description of activity taking place in the classroom and the evaluation of that activity. Several of the null hypotheses were not rejected, but there was enough doubt to warrant further investigation. The levels of the Dominant and Responsive Ratios were such that larger sample sizes may have made a difference.
3. There is a need to further integrate the planning of curriculum and lesson objectives with the actual presentation and evaluation of instruction. IVA should be redesigned to act as both a preassessment and an evaluation tool.

Concluding Remarks

With the problems facing education today, both real and imagined, there is a growing need to demonstrate the effectiveness of the teaching process. Test scores are just one avenue to measure effectiveness, another is the satisfaction students and parents express with the program. Programs need to be designed that foster active engagement of the learner in the process. Attitudes about education

need to change; everyone must view education as a necessary and worthwhile endeavor. Resistance to change within our present system must be removed and a commitment made to continuing process improvement. As part of this process improvement, the evaluation of teachers needs to improve through a procedure that provides the necessary feedback and a commitment to use the information to improve the entire system.

To this end, IVA can be used to improve the system in three ways: (1) to help faculty make midcourse corrections in their teaching, (2) to help with preteaching preparation, and (3) to assist in prehiring assessment of potential faculty. Of the three improvements just enumerated, using IVA as a prehiring assessment tool offers the greatest benefit. Using IVA to make predictions of success, potential faculty can be screened during assessment center presentations or lecture demonstrations. This affords institutions the ability to ensure excellence in teaching, which in turn will contribute to greater student satisfaction and permit elevated learning in the classroom.

REFERENCES

Abrami, P. C. and D. A. Mizener. "Does the Attitude Similarity of College Professors and Their Students Produce 'Bias' in Course Evaluations?" American Educational Research Journal, 20 (1983):123-136.

Abrami, P. C., R. P. Perry, and L. Leventhal. "The Relationship Between Student Personality Characteristics. Teacher Ratings, and Student Achievement." Journal of Educational Psychology 74 no.1 (1982): 111.

Beer, C. T. and G. G. Dardenwald. "Gender Differences in Adult Student Perception of College Classroom Social Environments." Adult Education Quarterly 40 no. 1 (Fall 1989): 40.

Brookfield, S. D. Understanding and Facilitating Adult Learning. San Francisco: Jossey-Bass Publishers, 1986, 9.

Cross, K. P. "The Adventures of Education in Wonderland: Implementing Education Reform." Phi Delta Kappan 68 no. 7 (1987): 496

Draper, N. R. and H. Smith. Applied Regression Analysis. New York, NY: John Wiley & Sons, Inc., 153.

Doyle, K. Student Evaluation of Instruction. Lexington, Massachusetts: Lexington Books, 1975, 19.

Ellett, C.D. "STAR: System for Teaching and Learing Assessment and Review (1990):

Feldman, K. A. "The Superior College Teacher From the Students' View." Research in Higher Education 5 (1976): 243.

Flanders, N. Analyzing Teaching Behavior Reading, Mass,: Addison-Wesley, 1970, 35.

Fenker, R. M. "The Evaluation of University Faculty and Administrators: A Case Study." Journal of Higher Education, XLVI:6 (November/December 1975)

Gaski, J. F. "Construct Validity of Measures of College Teaching Effectiveness." Journal of Educational Psychology 79, no. 3 (1987): 327.

Gilligan, C., In a Different Voice. Cambridge, MA: Harvard University Press, 1982, 23.

Glassman, R. B. "Course Evaluations: Are Half of Us Really 'Below Average?'" Academe 74:44 (JL/AG 1988): 11.

Gunter, B. "A Perspective on the Taguch Methods." Quality Progress (June 1987): 44.

Guskey, T. R. Improving Student Learning in College Classrooms. Springfield, IL: Charles C. Thomas, 1984, 15.

Handler, J. "Math Anxiety in Adult Learning." Adult Learning (April, 1990): 20-21.

Hoover, T. "An Experimental Study of a eomputer Assisted Teacher Training System Using Flanders' Interaction Analysis Category System Providing Immediate Feedback of Teaching Behavior to Naive Subjects." Ph.D diss., University of Nebraska, 1975.

Hoover, T. "Guidebook and Directions for a Computer Program Titled IVA." Unpublished Work in Progress, Loyola University of Chicago, 1989, 3.

Hutchings P. and T. Marchese. "Watching Assessment: Questions, Stories, Prospects." Change 22 (July/August 1990): 14.

Ishikawa, K. What is Total Quality Control? The Japanese Way. trans. David J. Lu. Englewood Cliffs, NJ: Prentice-all,Inc. 1985, 44.

Jones, J. "Students' Ratings of Teacher Personality and Teaching Competence." Higher Education 18 no. 5 (1989): 551.

Juran, J. M. and F. M. Gryna, Jr. Quality Planning and Analysis From Product Development Through Use. 2nd Ed. New York: McGraw-Hill, 1980, 1.

Kerlinger, F. N. Foundations of Behavioral Research. 3d ed., New York: Holt, Rinehart, and Winston, 1986, 405.

Loesch, T. and R. Foley. "Learning Preference Differences Among Adults in Traditional and Nontraditional Baccalaureate Programs." Adult Education Quarterly 38 no. 4 (Summer, 1988): 224.

Marsh, H. W. "Students' Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research." International Journal of Educational Research 11 no. 3 (1987): 275.

Marsh, H. W. "Students' Evaluation of College/University Teaching: A Description of Research and an Instrument." Presented at the Annual Meeting of the Australian Association of Research in Education, Sydney, Australia, November 1980, 7.

Marsh H. W.and J. U. Overall. "Validity of Students' Evaluations of Teaching: A Comparison with Instructor Self Evaluations by Teaching Assistants, Undergraduate Faculty, and Graduate Faculty." Paper presented at Annual Meeting of the American Education Research Association, San Francisco, 1979, ERIC Document Reproduction Service No. ED177205.

Measelle, R. L. and M. Egol. "A New System of Education: World-Class and Customer-Focused." St. Charles, IL: Arthur Anderson Consulting, 1990, 2.

Menges, R. J. and M. J. Kulieke. "Satisfaction and Dissatisfaction in the College Classroom." Higher Education 13 (1984): 255.

Murray, H. G., J. P. Rushton, and S. V. Paunonen. "Teacher Personality Traits and Student Instructional Ratings in Six Types of University Courses." Journal Educational Psychology 82 no. 2 (1990): 250.

Naftulin, D. H., J. E. Ware, and F. A. Donnelly. "The Doctor Fox Lecture: A Paradigm of Educational Seduction." Journal of Medical Education 78 (1973): 630.

Norusis, M. J. The SPSS Guide to Data Analysis. Chicago, IL: SPSS Inc., 1986, 346.

Ostmoe, P., H. Van Hoozer, A. Scheffel and C. Crovell. "Learning Style Preferences and Selection of Learning Strategies: Considerations and Implications for Nurse Educators." 23 no. 1 (1984): 27.

Perry, R. P., P. C. Abrami, and L. Leventhal. "Educational Seduction: The Effect of Instructor Expressiveness and Lecture Content on Student Ratings and Achievement." Journal of Educational Psychology 71 no. 1 (1979): 107.

Pierson, M. J. and S. B. Springer. "Can Anything Good Come from Non-traditional Degree Programs?" Lifelong Learning: An Omnibus of Practice and Research 11 no. 5 (1988): 20.

Pohlmann, J. T. "A Description of Effective College Teaching in Five Disciplines as Measured by Student Ratings." Research in Higher Education 4 (1976): 335.

Remmers, H. H. "Reliability and Halo Effect on High School and College Students' Judgements of Their Teachers." Journal of Applied Psychology 18 (1934): 620.

Ross, J. M. "Critical Teaching Behaviors as Perceived by Adult Undergraduates." Paper presented at the Annual Meeting of the American Education Research Association, March 27-31, 1989, 17, EDRS ED 311015, microfiche.

Rossmann, J. E. and E. El-Khawas. "Thinking About Assessment: Perspectives for Presidents and Chief Academic Officers." (Washington, D.C.: American Council on Education and the American Association for Higher Education, June 1987), 1.

Scriven, M. "Summative Teacher Evaluation" in J. Millman (ed.) Handbook of Teacher Evaluation. Beverly Hills: Sage Publications, 1981, 244, quoted in J. Jones, "Students' Ratings of Teacher Personality and Teaching Competence." Higher Education 18 (1989): 552.

The Ernst & Young Quality Improvement Consulting Group, Total Quality An Executives's Guide for the 1990s. Homewood, IL: Dow Jones-Irwin, 1990, 4.

Tollefson, N., J. S. Chen, and Kleinsasser. "The Relationship of Students' Attitudes About Effective Teaching" Educational and Psychological Measurement 49 (1989): 529.

Tollefson, N., H. Wigington, and P. McKnight. "Course Ratings as Measures of Instructional Effectiveness." Instructional Science 12 (1983): 389.

Truex, M. H. "Factors Critical to College Teaching Success or Failure." Improving College and University Teaching 23 no. 4 (1975): 236.

Williams, R. G. and J. E. Ware. "Validity of Student Ratings of Instruction Under Different Incentive Conditions: A Further Study of the Dr. Fox Effect." Journal of Educational Psychology 68 (1976): 48.

Donald R. Carter

Loyola University Chicago

USING INSTRUCTIONAL VERBAL ANALYSIS AS A PREDICTOR OF INSTRUCTIONAL PERFORMANCE IN A GRADUATE MANAGEMENT PROGRAM

The purpose of this study is to investigate the usefulness of using the Instructional Verbal Analysis (IVA) computer program as a means of predicting instructor performance in the classroom based on global ratings of the students. IVA is based on the work of Ned Flanders whose original study was designed to provide feedback to teachers to assist them in becoming more effective. Hoover has modified Flanders' work for a computer system to gather data on verbal activity. IVA in is present form has been further modified to take into account not only verbal behavior but presentation behaviors as well.

Twenty-four faculty members participated in this study. All of the subjects were part-time instructors with at least 10 years of business experience in the content area they were teaching. Each instructor underwent extensive preparation prior to the term in which they were teaching which. Due to the lack of formal teacher training their preparation was carefully monitored.

The research found that faculty performance is predictable using three variables: (1) enthusiasm, (2) rapport with the

students, and (3) organization of the presentation. The regression r square indicated that about 81% of the variance in student ratings of the faculty was accounted for by the three variables.

VITA

Donald R. Carter is presently Director of Institutional Research and curriculum coordinator of the Information Systems area for Keller Graduate School of Management, Chicago, Illinois. Prior to this position Don was an Associate Academic Dean for the West Suburban Center of Keller Graduate School. He has been with this organization for the past two years. From 1983 to 1989 Don was Director of Academic Computing for Columbia College in Chicago, Illinois. Before moving to the post secondary level Don taught high school data processing and accounting for Naperville Central High School and Oak Lawn Community High School.

Don received his Bachelors Segree from DePaul University in Business Education in 1973. He attended Northern Illinois University from 1975 through 1977 where he received a Masters of Science in Educational Administration, in 1983 he received a Master of Arts in Computer Science from DePaul University. Don received his Doctor of Philosophy in Curriculum and Instruction from Loyola University Chicago in May, 1992.

Don is currently a member of the Data Processing Managers Association and the Association of Computing Machinery. In addition to his interests in computer science, Don's current research is in effective instruction and quality.

APPROVAL SHEET

The dissertation submitted by Donald R. Carter has been read and approved by the following committee:

Dr. Todd Hoover, Director
Associate Professor, Curriculum and Instruction
Loyola University Chicago

Dr. Jack Kavanagh
Professor, Counseling and Educational Psychology
Loyola University Chicago

Dr. Ron Morgan
Professor, Counseling and Educational Psychology
Loyola University Chicago

Dr. Patrick Mayers
Vice President, Academic Affairs
Keller Graduate School of Management

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the Committee with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

April 15, 1992
Date

Todd Hoover
Director's Signature