

379
NB/d
NO. 3172

PERFORMANCE EVALUATION OF COMMUNITY COLLEGE MANAGEMENT
INSTRUCTORS USING STUDENT ACHIEVEMENT
AS THE CRITERION

DISSERTATION

Presented to the Graduate Council of the
University of North Texas in Partial
Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

James McKernon Jones, B.S., M.B.A.

Denton, Texas

May, 1990

Jones, James McKernon, Performance Evaluation of Community College Management Instructors Using Student Achievement As the Criterion. Doctor of Philosophy (Higher Education), May, 1990, 78 pp., 7 tables, bibliography, 51 titles.

This study concerns the relationship between student evaluation of instruction and student achievement in the field of management at the community college level. Purposes of the study were to determine the subjective student evaluation of instructor performance in introductory classes of management, student achievement in the class upon completion of the course, and the relationship between the student evaluation of instructor performance and student achievement in knowledge of the course.

The population studied was all 10 sections of the Principles of Management course taught by 8 instructors at Del Mar College in Corpus Christi, Texas during the fall semester of 1988. A pretest-posttest design was used to determine student achievement scores. The College Board provided sufficient copies of two versions of the College Level Examination Program (CLEP) tests for Introduction to Management for the pretest and posttest. A special statistical technique using multiple regression was used to calculate an achievement score for each student that was

adjusted for entry level knowledge. Student evaluations of instructor performance were paired with the achievement scores and grades students received from the instructor. Additional confidential demographic data was obtained about the students and the instructors.

Major findings of the study concluded there is no significant relationship between the student achievement scores and student evaluation of instructor performance. There was a wide variance in correlation of student grades and student achievement scores when individual sections or individual instructors were examined. The overall correlation of grades and achievement scores was statistically significant and was the highest of any of the factors studied. The study recommends using more objective measures of student achievement in evaluating faculty performance.

TABLE OF CONTENTS

	Page
LIST OF TABLES	v
CHAPTER 1	1
INTRODUCTION	1
Statement of the Problem	1
Purposes of the Study	2
Research Questions	2
Background and Significance	3
Performance Evaluation in the Teaching Profession	7
Significance of the Study	9
Limitations and Delimitations of the Study	9
Summary and Organization of the Study	10
CHAPTER II	13
REVIEW OF THE LITERATURE	13
Introduction	13
Performance Evaluation in General	13
Performance Evaluation of Faculty	15
Sources Used in Evaluation of Instruction	15
Methods Used in Evaluation of Instruction	16
Student Evaluation of Faculty	17
Student Achievement as a Measure of Faculty Effectiveness	18
Other Methods of Measuring Instructional Effectiveness	19
Summary of the Literature	20
CHAPTER III	25
METHODS AND PROCEDURES	25
Introduction	25
Research Setting and Population	25
Collection of Data	26
Pretest and Posttest Administration	28
Administration of the Student Evaluation of Instruction	29
Treatment of Data	30

Relationship Between Evaluation and Achievement in General	31
Correlation Between Student Achievement and Evaluation Classified by Instructor	31
Difference in Evaluations or Achievement Between Instructor's Sections	32
Difference Between Final Grade and Achievement . .	32
Summary	32
CHAPTER IV	35
RESEARCH FINDINGS	35
Introduction	35
Overall Student Evaluation of Instruction and Student Achievement Scores	35
Student Evaluation of Instruction and Student Achievement Scores by Instructor and Section	37
Student Evaluation of Instruction and Student Achievement Scores in Sections Taught by the Same Instructor	38
Relationship Between Final Student Grade, Student Evaluation of Instruction and Student Achievement Scores	40
Other Findings	41
Summary of Major Findings	45
CHAPTER V	47
SUMMARY, CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH	47
Introduction	47
Summary of Major Findings	48
Discussion of Findings	49
Conclusions	52
Recommendations for Further Study	53
APPENDIX	56
BIBLIOGRAPHY	74

LIST OF TABLES

Table	Page
1. Correlation of Student Evaluation of Instruction with Student Achievement Scores	36
2. Correlation of Student Evaluation of Instruction with Student Achievement Scores by Section and Instructor	37
3. Differences Between Sections Taught By Same Instructor Based Upon Student Achievement Scores and Upon Student Evaluation of Instruction . . .	39
4. Correlations By Sections Between Student Grades, Student Achievement Scores, and Evaluation of Instruction	40
5. Evaluation of the Instructors By Individual Students in Principles of Management	42
6. Differences Between Sections and Instructors in Student Achievement, Evaluation of Instruction, and Student Grade	43
7. Student Achievement Scores, Evaluation of Instruction and Student Grades by Section and Instructor	44

CHAPTER 1

INTRODUCTION

Performance appraisal of human resources is a valuable management tool that can be used to determine the degree of contribution an individual is making toward predetermined individual and organizational goals. It is used to determine compensation as well as aiding in individual development toward becoming more effective and efficient on the job.

A key to effective performance appraisal is the determination of the goals which individuals are expected to meet. Goals are difficult to measure, especially when the productive output of the individual is neither obvious nor quantifiable as in higher education. When individual goals are not clearly specified and integrated with organizational goals, communication about those goals and the measurement of their attainment becomes virtually impossible.

This study then focuses on the performance evaluation of community college instructors as representative of occupations or professions that are difficult to measure.

Statement of the Problem

This study concerns the relationship between student evaluation of instruction and student achievement in the field of management at the community college level.

Purposes of the Study

The purposes of the study are to determine;

1. Subjective student evaluation of instructor performance in introductory classes in management
2. Student achievement in a principles of management class upon completing the course.
3. The relationship between student evaluation of instructor performance and student achievement in knowledge of the subject.

Research Questions

1. Is there a significant relationship between student evaluation of management instructors and student achievement?
2. Is there a significant relationship between student evaluation of individual instructors and student achievement in individual classes?
3. Is there a significant difference in the relationship of student evaluation of the instructor and student gains in knowledge in different sections taught by the same instructor?

4. Is there a significant difference between the final grade given to the student, and the student evaluation of the instructor and student achievement?

Background and Significance

Performance evaluation is a critical management function that has a significant impact on all aspects of an organization as well as on the individual(s) being evaluated. It is essential for determining such things as the compensation and development of the individual.¹ To understand the problem, it is important to know what the purposes of performance evaluations are, how these may conflict, what types of performance evaluation are available and the controversy that exists over the "best" method.

There are several purposes for performance evaluation. First, performance evaluation is an important management control function for all organizations. In order to determine if and when individual and organizational objectives are being met, not met or exceeded, there must be ways to measure actual performance against expected standards. The standards that are established for the individuals must be congruent with the standards of the organization or else the organizational objectives will not be met.²

In addition to determining if organizational goals are being met, performance evaluations have two main purposes when dealing with individuals. These purposes include

administrative decisions such as salary level, promotions, raises, demotions, transfers and terminations. The other purpose is to determine what training and development needs exist for the individual to be of greater value to the organization and for their career advancement.

Two other reasons for performance evaluation should be recognized but will not be part of the major emphases of this study. They are the need to meet legal requirements and the use of performance evaluation for providing data to other parts of the management system such as for costing, scheduling, and budgeting.³

There seems to be common agreement that the more objective the performance evaluation, the more valuable is the appraisal. The disagreement comes when deciding what has to be measured objectively. Confusion over the use of the word "objective" contributes to the problem.⁴

The two purposes of compensation determination and training and development also represent major areas of conflict. Some experts believe that one performance evaluation can be used for both purposes. They feel the same performance evaluation should be used for promotions and raises as is used for development of the individual.⁵

Other authorities believe that individuals should be compensated based upon their direct objective contribution to organizational goals since they are so closely interwoven. Their personal development is a separate issue

and, if effective, simply results in increased productivity, which is then more highly compensated. They feel the two approaches to performance evaluation do not have a common ground and a separate evaluation should therefore be conducted for each purpose. Advocates of this approach believe that if employees expect a performance evaluation to determine their compensation, they will not openly discuss weaknesses that they need to overcome and how that will be accomplished. If a performance evaluation is done with the employee's understanding that it is to help them perform better but will not affect pay or status, chances are they will more readily accept discussion about personal improvement.⁶

Techniques of performance evaluation fall into two broad categories; objective or outcome based and, subjective or judgement based. Synonyms for the two techniques are as follows.

outcome	vs	input
results oriented	vs	behavior based
objective	vs	subjective
measurable	vs	judgmental

Objective performance evaluation measures the productive output of an individual. It is more commonly used to measure operating-level personnel because quantitative and qualitative measures are more readily available at that level and function.⁷ Examples of quantitative performance

evaluation include number of units produced, amount of sales, and value of production. The quality of the work generated can be determined by the scrap or waste generated, number of rejects, errors, or frequency of complaints. To evaluate performance, output quantity and/or quality are compared with some standard or goal. Standards are predominantly based on the performance of others doing the same thing or the individual's past performance.⁸

Objective measures of output are more difficult to determine when it comes to most professional and managerial positions. Some think it is highly impractical to search for the "ultimate" criteria where ultimate means the output.⁹ Subjective or behavior based performance evaluation is more commonly used with these people. Subjective performance evaluation is "where ratings are a direct function of human judgement".¹⁰ The standards that are established against which to measure performance are what some people believe are the traits, characteristics, actions or behaviors that are demonstrated by the person doing that job "effectively".

Recognizing the shortcomings in subjective evaluations, some elaborate quantitative systems and other programs have been devised to try to eliminate or reduce the subjective factor of human judgement and make them more objective. These methods range from variations of Management by Objectives, and Graphic Rating Scales, to Behaviorally Anchored Rating Scales.¹¹ None of these methods has been

accepted by all interested parties. These include the organization, supervisors, the general public, unions, regulatory agencies, and the individuals themselves. Something more concrete is needed as proof of performance. Carroll and Schneier, in their book Performance Appraisals and Review Systems (PAR) state, "A PAR system must . . . address what results are attained, as well as the methods used to attain them." ¹²

Performance Evaluation in the Teaching Profession

Teachers in general, and college teachers in particular, are subjected to performance evaluations that predominantly fall into the subjective or judgmental classification. Evaluations from students, peers, supervisors and self evaluation almost always concentrate on the evaluator's subjective judgement of how well the teacher does things that are considered to be what a good teacher does. Less emphasis is placed on what the student gains in knowledge or skills, presumably because student gains are harder to measure.¹³ In attempts to make subjective measurement techniques more objective than judgmental, elaborate quantification techniques have been developed, including systems where an attempt is made to determine which of the teaching characteristics are most valuable.¹⁴ Some characteristics that were assumed to be desirable

actually turned out to have a negative effect on academic accomplishment and/or student self esteem. ¹⁵

Attempts have been made to develop more objective performance standards for management personnel; and these have been used for the evaluation of professionals, such as teachers, as well. The best known of the upper level or management outcome related performance evaluation systems is Management By Objectives (MBO) in which managers and subordinates participate in the setting of objectives for a specific period. They are measured at predetermined checkpoints, and at the end of the period, to evaluate performance. In addition to being a performance evaluation system, MBO is also considered by some to be a management system, a management philosophy, and a planning and control mechanism. Originally intended to have objectives as defined in the program be as outcome related as possible, many applications of MBO have developed behavioral or activity based objectives. As a result there are many applications of MBO that have become very subjective in nature. Whether used in activity or results oriented environments, MBO is very good for developmental objectives but is useless for determining compensation and promotion. There really is no method that is considered good for the latter purpose, according to Bernardin and Beatty.¹⁶

A combination of the two methods is the recommended practice for effective performance evaluation. There is

still controversy over how much of each of the methods should be emphasized. Generally the more objectivity that the performance evaluation contains the better, as long as subjective evaluation is not eliminated from consideration. The subjective is generally less valid when it comes to measuring effective performance, but objective measures are usually the most difficult to develop.

Significance of the Study

Under the circumstances previously described, if objective criteria were established and accepted, outcome or results oriented goal attainment in an MBO environment would be extremely valuable in determining compensation and promotion decisions for teachers as well as other professionals. Justification for departmental programs and claims of institutional and organizational effectiveness would be easily documented and indisputable. Standards of comparison between all elements would be available for examination, evaluation and action.

Implications outside the academic field are obvious as well. If methods can be developed to utilize realistic objective outcome as a major part of performance evaluation, all areas where subjective performance evaluations are currently used would be subject to study to determine what objective results-oriented criteria could be developed for more effective performance evaluation.

Limitations and Delimitations of the Study

There were certain existent or planned factors in the study that could have an effect on the outcome. One was the Hawthorne effect that suggests students who were aware they were subjects of the study would perform differently than normal. Grade levels of other students taking the same course in other recent semesters showed no significant difference.

Delimitations of the study included the use of one community college and those students enrolled in the Principles of Management course. That course was used in the study because it had the highest enrollment of any course in the management programs and a College Level Examination Program test was available for that subject. Initial class size ranged from 15 to 30.

Summary and Organization of the Study

There is a need for re-examining the use of subjective performance evaluation. The controversy over the methods used have driven those who recognize the need for effective performance evaluation to use methods that seem to be objective but are only quantifying subjective measures, such as the emphasis in recent history on graphic rating scales and behaviorally anchored rating scales. Research is needed to find out if the performance appraisals, upon which so

much depends for the individual and the organization, are really accurate and doing the job they are supposed to do.

Chapter I presents the introduction and overview of the study and some of the background and the reasons why the study is significant. Chapter II contains some of the significant literature on this subject, pointing to some of the key areas mentioned in this chapter. Methods of research, the collection and the treatment of the data are described in Chapter III. The findings of the research are reported in Chapter IV. Finally, in Chapter V, the study is summarized and includes the conclusions that have been drawn from the major findings. Recommendations for further research are also presented.

CHAPTER BIBLIOGRAPHY

1. John Bernardin and Richard W. Beatty, Performance Appraisal: Assessing Human Behavior at Work, (Boston: Kent Publishing Co., 1984), 5
2. Ibid. 119
3. Stephen J. Carroll, and Craig E. Schneier, Performance Appraisal and Review Systems: The Identification, Measurement, and Development of Performance in Organizations, (Glenview, Il.: Scott, Foresman and Co., 1982) 130; Bernardin, 191
4. Carroll, 29
5. L.L. Cummings and Donald P. Schwab, Performance in Organizations, (Glenview, Il.: Scott, Foresman & Co., 1973), 91-95.
6. Bernardin, 234
7. Carroll, 29
8. Ibid. 131
9. Ibid. 30
10. Bernardin, 116
11. Craig Eric Schneier and Richard Beatty, "Combining BARS and MBO: Using an Appraisal System to Diagnose Performance Problems.", Personnel Administrator 24 (Sep. 1979): 51-60.
12. Carroll, 39; Craig Eric Schneier and Richard W. Beatty, "Developing Behaviorally-Anchored Rating Scales (BARS).", Personnel Administrator 24 (Aug 1979), 59-68.; Roger J. Plachy, "Appraisal Scales that Measure Performance Outcomes and Job Results," Personnel, (May-June 1983)
13. Robert S. Soar, Donald M. Medley and Homer Coker, "Teacher Evaluation: a Critique of Currently Used Methods.", Phi Delta Kappan, (Dec. 1983) 239.
14. Schneier, 51-60.; William J. Kearney, "Behaviorally Anchored Rating Scales--MBO's Missing Ingredient.", Personnel Journal, (Jan. 1979) 20-25.
15. Soar, 239.
16. Bernardin, 234.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

This chapter reviews the literature on performance evaluation, its purposes in general and how it applies to faculty, particularly in community colleges. Emphasis was placed on the methods and the sources of information used. Material reviewed included books that are current texts, classics in the field, and contemporary thinking on the subject. Professional journals and other publications, published and unpublished were also examined for additional thinking on the subject.

Performance Evaluation in General

According to Stephen J. Carroll, H. John Bernardin and Richard I. Miller, performance evaluation's major objectives are twofold.¹ One is to aid management in making personnel decisions for such administrative actions as raises, promotions, terminations and transfers. The other is for personal and professional development of the individual employee. Miller feels that one evaluation can be used for both purposes even though the objectives of the rater and

ratee may not be congruent because of selective perception, poor communication, conflicts of personality or other reasons.²

R. L. Taylor and R. A. Zawacki reported in 1984 that the graphic rating scale was the most popular appraisal form used in business.³

Behaviorally Anchored Rating Scales is the most popular form of graphic rating scale.⁴ Business and industry are going through a change according to C. Reinhardt in the use of performance appraisal.⁵ Programs and tools such as Management By Objectives (MBO) and Behaviorally Anchored Rating Scales (BARS) have been popular and are still the most widely used. However business is starting to avoid these collaborative and participative techniques and are moving toward computer based systems and assessment centers.

One of the reasons for the change is the lack of satisfaction in business with the results of current appraisal systems. Roger Plachy, Reinhardt, P. C. Smith and L. M. Kendall call attention to the subjectivity of raters as one of the major problems. Smith says, "No rating scale is really proof against distortion by a rater who really wants to do so."⁶ William I. Sauser, Jr. brings up the cost and time to develop acceptable criteria and train raters as being deterrents to effective evaluation.⁷ Dissatisfaction remains particularly high when personnel decisions are not

helped by the evaluation process. As John Bernardin says, "no method of appraisal has been a proven success for merit-pay or promotional decisions".⁸ There still is positive reaction to the use of performance evaluation for purposes of professional development provided the employee is motivated and the resources are made available to provide the development.⁹

Performance Evaluation of Faculty

Reinhardt reports that educational institutions are interested in what is being done in business, and they are using or modifying business systems for use in education.¹⁰ The problems are the same regardless of whether the organization is education or business according to Robert S. Soar, Donald M. Medley and Homer Coker.¹¹ The purposes of evaluation are the same; personnel decision making and professional development.¹²

To evaluate faculty, Lawrence H. Poole and Donald Dellow, list classroom effectiveness, advisement effectiveness, professional development and college/governance activities as the primary areas for consideration.¹³ All institutions ranked instruction as the most important factor according to a study of southern colleges and universities by James E. Boyd and E. F. Schietinger. Other factors vary in importance depending on the goal and purpose of the institution. ¹⁴

Sources Used in Evaluation of Instruction

Both Bernardin and Carroll list five potential sources of performance evaluation in general; peer, supervisor, external, subordinate and self.¹⁵ Students are given recognition as a separate source in academic institutions since 88% of institutions surveyed by Boyd and Schietinger reported doing some form of student evaluation of faculty.¹⁶ In the same survey, Boyd and Schietinger found that the superior was predominantly the source when advancement decisions were made and that the source for development decisions was also superiors but with equal attention to students.

Methods Used in Evaluation of Instruction

According to Richard I. Miller, procedures for evaluating classroom teaching are student evaluation, classroom visitation, teaching materials and procedures, special incident, and self appraisal.¹⁷ Another report by Peter Sedlin takes a different position. In a survey conducted in four year and upper level institutions, teaching competence impressions were primarily derived from the instructor's research and publication record. The presumption being if an instructor publishes a lot and well, then they must be good teachers in the classroom as well.¹⁸ Community colleges were not included in this survey. Faculty evaluation in community colleges presents a different

picture because of the de-emphasis of research and publication as compared with other institutions of higher education, according to William E. Cashin.¹⁹ Student achievement measures are another method of the evaluation of instruction that is recommended by P. A. Cohen, T. L. Wilson, H. R. Dodge and D. H. Mathews.²⁰

Tools that are used in faculty evaluation, at different times and with modifications, include graphic rating scales, (including Behaviorally Anchored Rating Scales), Management by Objectives, special checklists, narrative reports and observation.²¹ The most widely used is the graphic rating scale in various forms and developed by various sources.²²

Student Evaluation of Faculty

By far the most often mentioned method when faculty evaluation is discussed is student evaluation. Nearly every article and book on the subject comments either favorably or unfavorably about it. Oren Harari and Sheldon Zedeck used students for development of Behaviorally Anchored Rating Scales and consider them to be appropriate evaluators.²³ P. A. Cohen did a meta-analysis of validity studies and reported a strong validity of student ratings as measures of teaching effectiveness.²⁴ Richard K. Miller says, "Those who oppose the use of student appraisals deny the single most important data basis for judging teaching effectiveness", but he also points out that student

evaluations are the least likely factor to actually be used for performance appraisal.²⁵ Cashin points out that if student evaluations are to be used, there are many conditions regarding their effectiveness that need to be considered. He feels that student evaluation should be limited to evaluating classroom activity and what a student has learned.²⁶

Those opposed to the use of student evaluations are best summarized by Robert S. Soar, Donald M. Medley and Homer Coker when they wrote that ratings of teachers lack validity, are results of the halo effect, are useless and may be harmful.²⁷ An article by Coker, Medley and Soar reports on an extensive study in Georgia that found many of the traditionally accepted and recommended behaviors and characteristics of teachers, such as "making contact with a student who is not on task, and nonverbal communication", actually had a negative effect on student achievement and self esteem. Using praise or rewards, while relating positively to self-concept of the student, was negatively related to achievement of the student.²⁸

Many attempts have been made to validate student evaluations as indicators of teaching effectiveness, but for every one that shows a correlation with some measure there is another that contradicts.²⁹ Miller points out one of the major problems. No standards exist that are agreed to regarding criteria of teaching effectiveness.³⁰

Student Achievement as a Measure of Faculty Effectiveness

Cohen reports,³¹ student learning is the most important criterion of teaching effectiveness and there are few who disagree with such a statement. There are ways to measure student learning such as the student's final grade, the final exam or what the student expects in the way of a grade.³² Criticisms of using these measures include how different instructors grade and that such tests do not consider the difference in the students' knowledge at the beginning of the course.³³ Standardized tests in a pretest - posttest mode offset these criticisms but raise statistical problems of regression to the mean and the ceiling effect if the difference in the scores are used as an indication of the gain in student knowledge as a result of the course.³⁴ Finally there is always the criticism that no test can really determine the achievement of a student since intrinsic benefits cannot be measured by any test.³⁵ Those who agree with Cohen that student achievement can be a valid contributing factor to faculty evaluation include David L. Cook, the team of T. L. Wilson, H. R. Dodge and D. H. Mathews from Youngstown State University who borrowed the value added concept from business to measure student achievement, and the Coker, Medley, Soar group who did the research project of instructor characteristics vs. student achievement in the Georgia schools.³⁶

Other Methods of Measuring Instructional Effectiveness

The other methods of measuring effectiveness, including classroom visitation, examination of teaching materials and procedures, special incidents and self-evaluation, may or may not be used at a particular institution, may take on many different forms, but are not used individually as an exclusive measurement method. None of these methods has correlated in any significant manner with any important criteria of effective instruction.³⁷ One or more are used to supplement other methods to provide a more complete data base of information from which personnel and professional development decisions can be made.³⁸

Summary of the Literature

In summary a review of the literature discloses the most frequently mentioned part of faculty evaluation is evaluation of instruction. Student learning is the most important criteria of instruction. Student evaluations are the most common method of faculty instructional evaluation, aside from the personal observations of the faculty person's supervisor. Student evaluations are not used regularly in professional development and are considered rarely in decisions regarding personnel matters. Student evaluations are normally conducted using a modification of a graphic rating scale based on characteristics thought to be indicators of effective instruction. Student evaluations of

faculty have not been proven to correlate with student achievement. By comparison with research into subjective performance appraisal areas such as instructional characteristics, very little research has been done using student achievement as a criteria in faculty evaluation. What has been done has met with considerable criticism and resistance.

Performance evaluation is primarily the domain of the faculty person and the immediate supervisor. In addition the supervisor (or the institution) usually calls on input from students and sometimes from other sources, to obtain a more complete assessment for professional development purposes. When making personnel administration type decisions, the supervisor and the institution will use the same information but depend more on the supervisor's own personal judgement.

Educational institutions are just as dissatisfied with their performance evaluation systems as business and industry. They still use the systems because performance evaluation is essential and it is the best they can do. Educational institutions and industry are both continually striving to improve performance appraisal and are interested in sharing ideas and successes.

CHAPTER BIBLIOGRAPHY

1. Richard I Miller, Evaluating Faculty Performance, (San Francisco: Jossey-Bass, Inc., 1972), 111; John Bernardin, and Richard W. Beatty, Performance Appraisal: Assessing Human Behavior at Work, (Boston, Kent Publishing Co., 1984), 5; Stephen J. Carroll and Craig E. Schneier, Performance Appraisal and Review Systems: The Identification, Measurement, and Development of Performance in Organizations, (Glenview, Ill.: Scott, Foresman and Company, 1982) 5
2. Miller, 1972, 8
3. R. L. Taylor and R.A.Zawacki, "Trends in Performance Appraisal: Guidelines for Managers" Personnel Administrator, Vol. 29 No. 3, Mar. 1984, 71-80
4. Bernardin, 114
5. C. Reinhardt, "The State of Performance Appraisal: A Literature Review", Human Resource Planning, Vol 8 No. 2:104 October 1985.
6. P.C. Smith, and L.M. Kendall, "Retranslation of Expectations: An approach to the construction of unambiguous anchors for rating scales.", Journal of Applied Psychology, Vol 47, 1963, 149-155
7. William I. Sauser, Jr., "A Comparison of the Effects of Rater Training and Participation on Sources of Variance in a Set of BARS Ratings". Paper presented at the 26th Annual Meeting of the Southeastern Psychological Association, Washington, DC, March 26-29, 1980
8. Bernardin, 234
9. Miller, 1972, 4-12; Bernardin, 234
10. Reinhardt, 104
11. Robert S. Soar, Donald M. Medley and Homer Coker, "Teacher Evaluation: a Critique of Currently Used Methods.", Phi Delta Kappan, Dec. 1983, 239(13)
12. Miller, 1972, 11
13. Al Smith, ed.; Evaluating Faculty and Staff: New Directions for Community College Series #41, Lawrence H. Poole and Donald Dellow, Chapter 2, (San Francisco: Jossey-Bass, 1983) 19

14. James E. Boyd, and E. F. Schietinger, Faculty Evaluation Procedures in Southern Colleges and Universities, (Atlanta: Southern Regional Education Board, 1976) 78
15. Bernardin, 7
16. Boyd, 88
17. Miller, 1972, 26
18. Peter Sedlin, Changing Practices in Faculty Evaluation, (San Francisco: Jossey-Bass Publishers, 1984), 46
19. Al Smith, ed., Evaluating Faculty and Staff: New Directions for Community College Series #41, William E. Cashin, Chapter 5, (San Francisco: Jossey-Bass, 1983), 61
20. P. A. Cohen, "Student Ratings of Instruction and Student Achievement: A Meta-analysis of Multisection Validity Studies", Review of Educational Research, Vol 51, 1981, 283; T. L. Wilson, H. R. Dodge and D. H. Mathews, "Value Added Testing Is Likely To Expand", Marketing News, July 3, 1987, 25
21. Bernardin, 114,119; Carroll, 102; Robert J. Beebe, Use of the Behaviorally Anchored Rating Scale in Evaluating Teacher Performance. Paper presented at the 6th Annual Conference of the Evaluation Network Memphis, TN. September 30 - October 2, 1980
22. Beebe; Oren Harari and Sheldon Zedeck, "Development of Behaviorally Anchored Scales for the Evaluation of Faculty Teaching.", Journal of Applied Psychology, Vol. 58, No. 2, 1973, 261-265; Brian J. Kauffman and Joseph Madden, "The Development of Behaviorally Anchored Rating Scales for Student Evaluation of College Teaching Effectiveness." Paper presented at the 51st Annual Meeting of the Eastern Psychological Association, Hartford, CT, April 9-12, 1980; Taylor, 78
23. Harari, 263
24. Cohen, 1981, 281-309
25. Miller, 1974, 3,24
26. Smith, Cashin chapter 5, 61
27. Soar, 239
28. Homer Coker, Donald M. Medley and Robert Soar, "How Valid Are Expert Opinions About Effective Teaching?", Phi Delta Kappan, October, 1980 131-135
29. Cashin, 57

30. Miller, 1972, 24
31. Cohen, 283
32. Cohen, 300
33. Schuyler W. Huck, William H. Cormier and William G. Bounds, Jr., Reading Statistics and Research, (New York: Harper and Row, 1974), 133-147,247; Cohen, 306
34. Soar, 241
35. Miller, 1972, 8
36. David L. Cook, An Objective Component for the Evaluation of Teaching, Ed.D dissertation, (University of Tennessee), 1985; Wilson, 25; Homer Coker, Donald M. Medley and Robert Soar, "How Valid Are Expert Opinions About Effective Teaching?", Phi Delta Kappan, October, 1980, 131-135
37. Miller, 1972, 24; Coker, 133; Reinhardt, 105
38. Siegel, 255; Taylor, 75

CHAPTER III

METHODS AND PROCEDURES

Introduction

Chapter III contains detailed information about the methods and procedures of this study. To determine the relationship between student evaluation of instruction and student achievement in the field of management a set of research questions were developed to guide the collection and treatment of the data. The methods of collecting the necessary data and its treatment are described as well as the instruments used, the population utilized, and the demographics of this population.

Research Setting and Population

Del Mar College is a state and locally supported community college in south Texas. The college offers various programs including several in business.

The population of this study is all of the students enrolled in the Principles of Management course in the fall of 1988 who furnish usable data. Most of the 251 students who originally enrolled in the course were freshmen or sophomores; however a few who already hold degrees take the

course. Of the 251 students who originally enrolled, 171 took the final out of which 155 provided usable data. Therefore the population of this study is 155 student subjects.

The population included both day and night students, majors in two year management degree programs, majors in other degree programs where the course is a requirement, and a miscellaneous category of students who took the course for a variety of reasons. The Principles of Management course was selected because it has the largest enrollment of any one course in the management program and because of the availability of the nationally recognized standardized test with proven reliability and validity described below.

Approval, endorsement, and support from the College and the Department of Business Administration were secured. Faculty were advised of the study and were impressed with the necessity for their cooperation as the results of the research was deemed to be of value to the department, the college, and the individual faculty members themselves. All faculty were informed of the study in a meeting and received a memo prior to each of the tests giving complete instructions for their students and information for themselves.

Collection of Data

Two types of basic data were collected. One was the measure of student achievement in the course and the other was the evaluation of the instructor of the course provided by the individual students. Other data collected included demographics on the students and instructors, class section, and grade awarded by the instructor.

To measure student achievement, an internationally recognized and accepted standardized test was used for a pretest and posttest. Permission was obtained from the Program Director of the College Level Examination Program (CLEP) for the use of the Introduction to Management test for this study. Two separate and distinct versions of the CLEP test were used for the pretest and posttest. Both versions have been validated and the appropriate adjustment factors determined by CLEP. (see appendix A for a copy of the technical specifications). Raw scores can be converted according to the statistical table developed and provided by the College Level Examination Program. Results of the converted scores can then be compared with other tests on the same subject. The agreement calls for providing the Educational Testing Service with a copy of the results of the study.

A standard graphic rating scale tested and adopted by Del Mar College was used for the measurement of student

evaluation of instructors. This is a seven point scale with seven being the highest rating, one being lowest and zero representing no knowledge or comment by the student.

Seventeen items are answered, sixteen of which pertain to the instructor and the seventeenth pertains to the adequacy of the instrument. (see Appendix B for a sample of the form used for the student evaluation of the instructor)

Pretest and Posttest Administration

Both the pretest and the posttest were administered in each section of the course by either the researcher or the chairperson of the Business Administration department. This provided standardized administration and reduced or eliminated the influence of the individual instructor. Every student was given a memo prior to the pretest explaining the purpose of the test. (see appendix C for sample of memo.) Both tests were machine scored so that results were available to the instructors immediately. This was important particularly for the final exam. Scoring was done by the two people who administered the tests.

The pretest was administered during the first week of the semester. Students were advised that the standardized test would be all or part of their final exam grade and, although the pretest would not be part of their grade, it would help them prepare for the final.

The posttest was administered in the normal time allocated for final exams for the course. It was in addition to, or in substitution for, any final exam the instructor wanted to administer. Instructors were told to include the results of the posttest as all, or part of, the input for the final exam grade and to communicate that to their students. Prior to the exam, students were given a memo of explanation and a data sheet to fill out. The data sheet was returned with the test. (see appendix D for a sample of the memo and data sheet).

Administration of the Student Evaluation of Instruction

The evaluation of instructors by the students was completed at the same time as the administration of the final exam. The process was the same as the process used by the College with one exception. The evaluation was administered in the classes by the same two people who administered the pretest and posttest. Students filled out the evaluation sheets marking their numerical ratings as well as the comments. They transferred their scores to a computer card for the normal statistical tabulation by the college. For purposes of this study, each student was given a randomly generated code number which was recorded on a corner of the evaluation sheet. Once the scores were recorded and matched with the pretest, posttest and demographic data, the number was cut from the corner of the sheet and sent forward for processing by the college in the

normal way. This one exception to the usually totally anonymous method of evaluation was done to give more statistical power to the study by matching evaluation scores with the test scores.

During the evaluation process the instructor was not present in the class and had no access at any time to the evaluation materials. The results of the evaluation were provided to the instructor, as usual, after the semester was over and grades were reported.

Treatment of Data

Data was analyzed using causal-comparative methods and correlational analysis. The causal-comparative methods were used to determine the statistical significance of the data. The correlational analysis was used to determine the degree of the relationship that existed between variables.

Before any analysis could be done, two adjustments and intermediary calculations had to be made. First a method had to be developed to measure student achievement. To avoid the problems and criticisms inherent in measuring student gain,¹ a special method was developed to offset the effect of the incoming ability of the student as indicated by the pretest score. Using the pretest and posttest scores of all students in a multiple regression model, a regression equation was developed from which a predicted grade for each student was computed. This was the posttest grade the student would have

received based on a projection of their pretest score. The predicted score was adjusted by the residual of the regression. This adjusted posttest score allowed for the entry level knowledge the student had at the beginning of the course. The adjusted posttest could then be used as a measure of achievement with this particular population and for this particular purpose.

The other calculation concerned the student evaluation of the instructor. An attempt was made to apply each of the 16 items of the student evaluation separately as individual variables. The number of variables was too large for the size sample involved. The statistical analysis lost its strength and became misleading and meaningless using that many variables. Since Del Mar College uses the average score of the evaluation for personnel decisions, the average of the evaluations was used as the variable for this study.

Relationship Between Evaluation and Achievement in General

To study the first research question concerning the relationship between student evaluations of instructors and student achievement, a simple Pearson's Product Moment correlation was done in addition to the multiple regression mentioned above. Both of these methods gave indications of the relationship of the variables.

Correlation Between Student Achievement and Evaluation Classified by Instructor

For the second research question about the relationship between the student achievement and the evaluation classified by instructor, analysis of variance was done first to determine if there was a significant difference between instructors for each of the variables. As an additional check, Pearson's Product Moment correlation was used separately for each instructor to determine if there was any relationship.

Difference in Evaluations or Achievement Between Instructor's Sections

In the third research question, analysis of variance was used to determine if there was a significant difference between student achievement or instructor evaluation when broken down between different sections taught by the same instructor. There were two instructors who taught more than one section of Principles of Management and the analysis was made on just those two instructors' sections as well.

Difference Between Final Grade and Achievement

Final grades that the students received from their instructors were the subject of the last research question and were in letter grade form. These were recoded using a five point system where A = 5, B = 4, C = 3, D = 2, and F = 1. Recoding was simply for application in a computerized

statistical program and had no relationship to normal grade point calculations. Analysis of variance was used to determine if there was a significant difference between the grades and the student achievement by section of the class and by instructor.

Summary

In summary, the collection of data was for the purpose of obtaining measures of student achievement and the students' evaluation of their instructor as well as the final grade awarded by the instructor to the student. Other information such as the section of the class in which the student participated and the instructor for that section was collected. The treatments applied to analyze the data were primarily descriptive statistics with more complex models of multiple regression, analysis of variance and multiple correlational techniques. Multiple regression was used to calculate student achievement in a way that overcame problems in determining an accurate measurement for student achievement.

CHAPTER BIBLIOGRAPHY

1. Al Smith, and Lawrence H. Poole, Donald Dellow, Richard Behrendt, William E. Cashin, Evaluating Faculty and Staff: New Directions for Community College Series #41, (San Francisco: Jossey-Bass, 1983) 61; P A Cohen, "Student Ratings of Instruction and Student Achievement: A Meta-analysis of Multisection Validity Studies", Review of Educational Research, 1981 vol 51, 283

CHAPTER IV

RESEARCH FINDINGS

Introduction

This chapter reports the analysis of data that was collected and treated in an attempt to discover the relationship between student evaluation of instruction and student achievement. Each of the research questions is analyzed using the collected data. Other findings of interest developing from the study are reported and the summary of the major findings of the chapter are listed.

Overall Student Evaluation of Instruction and Student Achievement Scores

Research question one was established to seek data to determine the relationship between student evaluation of management instructors and student achievement scores. The relationship involving all management instructors and all sections of the introductory class in management is the focus of this research question. Table 1 presents data concerning correlation between the subjective student evaluation of instructor performance and student performance in the introductory management courses. The data in Table 1 show the correlation between the 192 students evaluations of

the 8 instructors and the achievement scores of these 192 students.

Table 1 - Correlation of Student Evaluation of Instruction with Student Achievement Scores

Number of Instr.	Number of Sect.	Number of Stdts	Student Instructor Evaluation		Student Achievement Scores		Correlation
			M	SD	M	SD	
8	10	N 192	5.6	1.35	46	9.67	.04

The overall correlation in Table 1 between student achievement scores and student evaluation of instruction was extremely low (-.044). Several factors might account for this lack of correlation. One factor may be the skewing of the student evaluations of instruction toward the high scores (7 being the highest possible). No evaluations were below 2. Another factor could be the evaluations are done before the grades are known reflecting possible suspicion by the students, despite the reassurance of anonymity, that the evaluations are available to the instructor before the grade is awarded. A third factor may be the connection between the feelings individual students may have about the ability of the instructor and the effect of these feelings upon their achievement scores.

Student Evaluation of Instruction and Student Achievement
Scores by Instructor and Section

The second research question attempts to secure data to assist in discovering the relationship between student evaluation of instruction and student achievement scores based on individual instructors and individual sections of the course. Table 2 contains data referring to the separate sections of the management course and to the correlation of instructor evaluations with student achievement scores.

Table 2 - Correlation of Student Evaluation of Instruction with Student Achievement Scores by Section and Instructor

Section	Instructor	Student Evaluation of Instruction		Student Achievement Scores		Correlation
		Mean	SD	Mean	SD	
0	B	6.62	.29	47.60	8.45	-.122
1	1*	4.84	1.39	45.22	8.37	.36
2	2*	4.76	.81	46.65	10.25	-.319
3	3	5.91	1.13	39.66	7.69	-.277
4	1*	5.74	1.06	46.84	9.48	-.264
5	4	6.61	2.43	46.85	11.82	.52
6	5	5.03	.60	45.57	9.13	.135
7	6	5.96	.86	46.97	10.55	.01
8	7	6.11	.29	48.42	8.45	-.312
9**	2*	4.78	1.78	51.92	7.90	.302
All Sections		5.60	1.16	46.00	9.67	-.044

* Instructors 1 and 2 both taught two sections each.

** Section 9 was a telecourse.

Viewing the data in Table 2 regarding student evaluations of instructors in individual sections and student achievement scores, the wide variation in

correlations seems to indicate that a pattern of relationship between student evaluation of instruction and student achievement scores is difficult to determine. Correlations vary between .52 and $-.31$. One section (section 7) shows almost no correlation.

It might be noted, however, that each section and its instructor can be viewed separately. For example in section 3, instructor 3 was rated somewhat high while class achievement scores were the lowest of all sections. Note section 9 where the student achievement score was high and the evaluation was low; however the highest achievement score was in this section -- the telecourse -- where there was limited interaction between students and instructor.

Student Evaluation of Instruction and Student Achievement Scores in Sections Taught by the Same Instructor

The third research question sought data to determine any differences in student evaluation of instruction and in student achievement scores in sections taught by the same instructor. Table 3 contains the student evaluations of instruction and student achievement scores for sections taught by the same instructor. Two instructors taught two sections each.

Table 3 - Differences Between Sections Taught By Same Instructor Based Upon Student Achievement Scores and Upon Student Evaluation of Instruction

Sect.	Instr.	Number of Students	Instructor Evaluation by Student			Student Achievement Scores		
			Mean	SD	F	Mean	SD	F
1	1	23	4.84	1.39	6.93*	45.22	8.37	.32
4	1	31	5.74	1.06		46.84	9.48	
2	2	12	4.76	.81	0	46.65	10.25	1.88
9	2	25	4.78	1.78		51.92	7.90	

* $p < .05$

The data in Table 3 indicate that between instructors teaching more than one section there is no significant difference in the mean student achievement scores. However it can be seen that there is a significant difference in the student evaluation of instruction between the two sections taught by instructor 1. A very small difference in the mean evaluation (based on a 7 point scale) can result in a significant difference.

It should be noted that section 9 was the telecourse, and this section had the highest student achievement scores. The on-campus sections were fairly close in the mean student achievement scores. Even with the high achievement scores in the telecourse, there was no significant difference in the means of the student achievement scores of the sections of either of the instructors. Although not presented in Table 3, the student evaluations of instruction were significantly

lower for instructor 2 than for instructor 1 ($F=4.03$ with $p<.05$).

Relationship Between Final Student Grade, Student Evaluation of Instruction and Student Achievement Scores

Table 4 shows the correlation between the major variables of final grades awarded to students by instructors, student achievement scores, and student evaluation of instructors. The wide variances in correlation of sections is shown by sectional breakdown.

Table 4 - Correlations By Sections Between Student Grades, Student Achievement Scores, and Evaluation of Instruction

Section	Std't Grades Mean	Student Achievement Scores Mean	Std't Evaluation of Instructor Mean	Final Grade and Achievement Score Correlation	Achievement Score and Evaluation of Instruction Correlation	Final Grade and Evaluation of Instruction Correlation
0	4.57	47.60	6.62	.16	-.12	.16
1	3.13	45.22	4.84	.77	.36	.59
2	3.58	46.65	4.76	.78	-.32	-.52
3	3.21	39.66	5.91	.26	-.28	-.22
4	2.81	46.84	5.74	.78	-.26	.00
5	3.70	46.85	6.61	.71	.52	.12
6	3.30	45.57	5.03	.82	.14	.50
7	3.94	46.97	5.96	.72	.01	-.32
8	3.64	48.42	6.11	.46	-.31	.43
9	3.12	51.92	4.78	.18	.30	-.43
All	3.349	46.00	5.60	.638	-.044	.112

Correlation between student achievement scores and the actual grade the student received from the instructor for

all students is moderately high, certainly the highest of any of the relationships in this study; and there are no individual negative correlations. There is a wide range of correlation values by section; however a high correlation should indicate that whatever grading system an instructor is using accurately evaluates the true achievement of the student in the subject. Conversely, a low correlation may imply that the grades the student received in that section were not a valid indication of the achievement of those students.

The relationship between evaluation of instruction and the grade the student received is low to non-existent. However a wide variance exists for individual sections. The range is even greater than the correlation by sections between evaluation and achievement.

As previously reported in Table 2 there is little relationship between evaluation of instruction and achievement scores. There is a wide variance when individual sections are examined and in some sections there are moderately negative correlations that could indicate low achieving students rating instructors high while high achievers rated their instructors low.

Other Findings

To indicate the general rating of instructors by these students, Table 5 shows the distribution of the student evaluations of instructors completed by all of the students in all of the sections of the course. Instruction is rated on 16 items by students using a scale of 1 to 7 with 7 being the highest score possible.

Table 5 - Evaluation of the Instructors By Individual Students in Principles of Management

Average Evaluation	Number of Students Giving Rating
6 to 7	75
5 to 6	41
4 to 5	32
3 to 4	22
2 to 3	2
1 to 2	0

There is a definite skewing of the evaluations toward the upper limits of the scale. This fact may limit the use of the evaluations as a performance evaluation tool because of the difficulty of differentiating at the upper end of the scale.

Student ratings which are in the upper range might suggest their satisfaction with instruction or their unwillingness to be critical. The uses of evaluation which does not discriminate therefore are limited.

In addition evaluations are administered to students while they are still in class before they receive their final grade. It is possible that this could affect the rating in spite of the great care that is taken to guarantee anonymity.

Using the data from all sections, Table 6 shows the significance of the difference between sections in the means of each of the measures of achievement scores, evaluation of instruction and grade. The table also shows the differences between instructors.

Table 6 - Differences Between Sections and Instructors in Student Achievement, Evaluation of Instruction, and Student Grade

	Student Achievement Scores	Student Evaluation of Instruction	Student Final Grade
Between Sections F Factor Significance	1.75 *p<.10	8.86 *p<.005	2.65 *p<.01
Between Instructors F Factor Significance	1.96 *p<.10	9.35 *p<.01	3.07 *p<.005

Note: Calculations were based on 10 course sections and 8 instructors.

Overall, the differences between sections are significant at a probability of .01 or higher in the final grades given and the evaluation of instruction by students.

The difference in achievement scores by students should not be considered as significant.

The difference between the grades given by instructor and the achievement scores may be indicative of other input factors such as personality, likes and dislikes, hard versus easy grading, and differences in grading tools. Students may not be rating instruction according to their achievement scores as was also indicated by the lack of correlation between achievement scores and evaluation in Table 4.

Table 7 displays the information in Table 2 with the addition of student grades by sections and instructors.

Table 7 - Student Achievement Scores, Evaluation of Instruction and Student Grades by Section and Instructor

Sec. No.	Instr. No.	Stdts N	Student Achievement Scores		Student Evaluation of Instruction		Students Grades	
			Mean	SD	Mean	SD	Mean	SD
0	8	7	47.60	8.45	6.62	.29	4.57	.29
1	1	23	45.22	8.37	4.84	1.39	3.13	1.39
2	2	12	46.65	10.25	4.76	.81	3.58	.81
3	3	24	39.66	7.69	5.91	1.13	3.21	1.13
4	1	31	46.84	9.48	5.74	1.06	2.81	1.69
5	4	20	46.85	11.82	6.61	2.43	3.70	1.06
6	5	20	45.57	9.13	5.03	.60	3.30	2.43
7	6	16	46.97	10.55	5.96	.86	3.94	.60
8	7	14	48.42	8.45	6.11	.29	3.64	.86
9	2	25	51.92	7.90	4.78	1.78	3.12	1.78
All Sections		192	46.00	9.67	5.60	1.16	3.35	1.16

As can be noted in Table 7, section 9 with instructor 2 had the highest mean achievement score but received the next

to lowest average grade given by the instructor. Section 9 was the telecourse with minimal instructor contact. The on-campus section, number 2, received a higher mean grade from instructor 2 but had a lower average achievement score. This may be explained by the personal relationship which might have been established on campus.

One instructor (instructor 8) received the highest evaluation and gave the highest grade. This might be explained by the fact that this section had only 7 students and was held on a military base. The fact that achievement scores are only in the average range does not substantiate the high evaluation and grades.

Summary of Major Findings

The following is a summary of the major findings of this study.

1. There is low correlation between the evaluation of instructors done by students and the achievement scores of the students considering all sections and all instructors.

2. There is a wide variation in correlations between student achievement scores and student evaluation of instruction considering individual instructors and sections.

3. While there are some differences in achievement scores when broken down by class sections and instructors, the differences were not statistically significant.

4. Concerning instructors teaching more than one section there is no significant difference in student achievement scores between sections; however there is a significant difference in evaluation of instruction between sections.

5. The highest correlation between any two variables is between student achievement and the grade the student received from the instructor, analyzing the figures of all the sections combined. The correlation of individual sections and instructors varies widely from high to low.

6. There is little correlation between the student evaluation of the instructor and the grade received by the student from the instructor. This applies in the overall statistical analysis as well as by sections and instructors.

7. There are significant differences between sections and instructors, in both student evaluation of instruction and the grades students received. These include sections taught by the same instructor.

CHAPTER V

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

Introduction

This study focused on the performance evaluation of community college instructors and examined the relationship between student evaluation of instruction and student achievement scores in the subject of management at the community college level. Investigated and analyzed were the relationships between student evaluation of management instructors, student achievement scores and the grades students received. The differences between instructors and individual sections of the course were also studied.

Methods and procedures included utilizing the scores on a nationally accepted standardized test to determine the achievement level of the student in the subject matter. The results were coupled with a linear regression technique to correct for variances of subject knowledge prior to the course. Other data collected for this study included the confidential evaluation of instruction by the student and the grade the student received from the instructor for the course. Various multiple regression and correlation techniques as well as descriptive statistics were applied in the study.

Summary of Major Findings

The following is a summary of the major findings of the study.

1. There is little relationship between the evaluation of instructors done by students and the achievement scores of the students considering all sections and all instructors.

2. There is a wide variation in correlations between student achievement scores and student evaluation of instruction considering individual instructors and sections.

3. While there are some differences in achievement scores when broken down by class sections and instructors, the differences were not statistically significant.

4. There is no significant difference in student achievement scores between sections taught by the same instructor; however there is a significant difference in evaluation of instruction between sections.

5. The highest correlation between any two variables is between student achievement scores and the grade the student received from the instructor, for all sections combined. The correlation of individual sections and instructors varies widely from high to low.

6. There is little correlation between the student evaluation of the instructor and the grade received by the student from the instructor for all sections combined or for individual sections and instructors.

7. There are significant differences between sections and instructors, in both student evaluation of instruction and the grades students received. These include sections taught by the same instructor.

Discussion of Findings

There are several reasons why little relationship might exist between student evaluation of instruction and student achievement scores. It is possible that the student achievement scores are not valid or reliable indicators of student achievement because of the method used to calculate them, even though the method used was recommended by authorities in the field, according to the literature,. Not knowing what the final grade might be, or concern about the confidentiality of the evaluation of instruction, might affect the objectivity of the student when completing the evaluation of instruction. The evaluation of instruction might not be a reliable indicator of what students actually achieve in the course and may be simply an indicator of personal popularity or lack thereof.

The same points can be made for the little relationship that exists between the grades the students receive from the instructor and the student evaluation of instruction, except that in this case the grade awarded by the instructor rather than a standardized test is used to determine student achievement. Grades are entered on the student permanent

record and therefore need to be an accurate measure of the student achievement in the course. Other variables could influence the limited relationship reported such as the objectivity of the instructor assessment of student achievement and the effect of the methods of grading. There are studies that report a stronger correlation between student evaluation of instruction and grades received from the instructor than what is indicated in this study.

The higher - and significant - correlation between the achievement scores of the students and the student grades can give support to the reliability and validity of both measures of student achievement. If the correlation were higher an even greater reliability for both measures could be assumed, and this would be worth exploring in further studies. If the method used to obtain student achievement scores is the more accurate measure, then the grades awarded by instructors need to be examined in an attempt to improve their accuracy. If the grade awarded by the instructor is more accurate, grades could be used with greater confidence as part of the assessment of instructor performance, with some controls to assure their continued accuracy. Research for additional methods of measuring student achievement, other than grades awarded by the instructor, could serve as the control mechanism mentioned previously.

The lack of a statistically significant difference in the student achievement scores between sections of the

course and between instructors can mean several things. The accuracy of the method used in computing the score could affect the outcome but this does not seem to be the case, as was pointed out previously in this section. Since there were some differences in the achievement scores of the different sections, the sample may not have been large enough to bring out the difference statistically. Another alternative could be that the students are learning about the same in each section on the average regardless of the personal influence of the instructor or the methods of instruction.

There were significant differences between sections of the course and between instructors in the student evaluations of instruction and in the grades given to the students by the instructor. This might be easier to analyze if there were some relationship between the two variables in general, or by section or instructor. If so, some assumptions could be made that the students were effectively evaluating the instructor based on their achievement level for the course. However, since this is not the case, and the student evaluations of instruction were skewed toward the high end of the scale, students may have given high ratings on the evaluation of instruction anticipating they might receive a higher grade by doing so. Another explanation is that the criteria in the evaluation of instruction are not reliable or valid indicators of good instruction. Since the average grades have significant differences between sections

and instructors as well, instructors may not be using effective methods to determine appropriate grades. For example with one instructor the grades awarded on the average were the inverse of the student achievement scores in the two sections the instructor was teaching.

Conclusions

The following are conclusions drawn from the study and the analysis.

1. There does not appear to be any factor normally considered in performance evaluation of faculty that relates to the actual student achievement. On the average students appear to be achieving at the same level in the course regardless of what section they are in. An instructor who is meeting or exceeding the personality and behavioral characteristics of what is supposed to be good instructional practice, does not necessarily produce higher achievement scores in students.

2. There does not appear to be any causal relationship between the student evaluation of the instructor and student achievement scores.

3. The influence of the instructor may not be as great as other factors that affect student achievement scores. Some possible factors reported in the literature that might have a more significant influence on the achievement scores of the student are extrinsic and

intrinsic values, family influences, work environment, personal relationships, finances, social and cultural influences, peer pressure and immediate critical incidents.

4. There appear to be additional factors involved in the final grade awarded by the instructor other than student achievement in the course.

5. Student grades may not be accurate indicators of student achievement as measured by standardized tests. While there was a moderate correlation between the two variables overall, the correlation between the variables in the individual sections varied widely.

6. Faculty evaluation should include an element that measures student achievement as accurately as possible. Including student achievement scores as an element in faculty evaluation could satisfy those who require a measure of accountability. Student achievement scores could also serve as an important factor for personnel decisions based on merit and provide a method to determine the need for professional development particularly for faculty who need evidence that such a need exists.

Recommendations for Further Study

Recommendations for further study call for continuing and expanding, but not replicating, this study. The new questions raised as a result of this study should be researched and analyzed. The recommendations are as follows.

1. The most significant need for further research as a result of this study is for continued development of new and better measures of student achievement. The one developed for this study was the best according to all the resources available. The application was appropriate for this study. Expanding to a wider scale among other students taking an introductory course in management principles would be necessary to determine if the results of this study continue to be reliable and valid.

2. Further research is needed for other quantitative measures of faculty effectiveness that relate to output. The relationship of grades awarded by the instructor to the achievement of the student measured by standard means might be one. This might be a measure of the effectiveness of the grading system used by the instructor as related to the true achievement of the student considering objectivity, subjectivity, and overall accuracy.

3. Research into other methods of measuring student achievement would be helpful to reassure faculty that accurate measurement is possible and the objective of improving student achievement is a realistic and measurable goal in performance evaluation.

4. Additional research is needed to determine if a significant difference continues to exist between the grades instructors give students who attain the same level of achievement. If so, either the achievement measurement is

not reliable or valid, or instructor grading needs further examination and possible corrective action.

5. The use of other standardized CLEP tests, such as for marketing principles and accounting in the business administration curriculum, should be tested to determine if results are comparable to the introduction to management test results. If the results are similar it would confirm their use as measures of achievement.

6. The telecourse had the highest student achievement of any section of the course. Because of the limited student contact with the instructor in a telecourse as compared with traditional classes, further study of this and other distance learning courses is necessary to determine if this is typical and, if so, why.

APPENDIX

APPENDIX A

COLLEGE LEVEL EQUIVALENCY PROGRAM (CLEP) SPECIFICATIONS BROCHURE

CLEP Introduction to Management

The information in this Guide is presented in two parts. Part 1 provides a description of the test and its development as well as sample test questions. Part 2 presents technical data on the test and will be of special interest to those with a background in tests and measurements.

College faculty and others who wish to determine the appropriateness of the College-Level Examination Program (CLEP) Introduction to Management test for awarding credit for or exemption from the introductory management course offered by the institution should find the information contained in this Guide helpful. Other program publications which provide additional information about the CLEP program are available from CLEP, CN 6601, Princeton, NJ 08541-6601.

Each institution has the opportunity to determine its own standards for granting credit. CLEP will help with suggested procedures and provide test books if a college wants to conduct a local standard-setting administration. The standard for granting credit that has been proposed by the American Council on Education (ACE) is the average test score earned by students in the reference group described on page 7 whose final grade in an introductory management course was C. The CLEP Introduction to Management test was designed to cover the instructional content offered in a one-semester course, and the minimum score recommended by the ACE for awarding credit is a scaled score of 47. (See Table 1 on page 6.)

I. The Test

Test Development History

The Introduction to Management test was initially developed in 1970 and was originally titled Introduction to Business Management. The 1978-79 Test Development Committee felt that the word "Business" should be dropped from the title in order to broaden the test to include application of management principles to nonprofit and governmental organizations. During the spring of 1979, two new forms of Introduction to Management were administered for the purpose of rescaling and re-norming because of substantial changes in test specifications.

The Test Development Committee for the current Introduction to Management forms was made up of the following college professors. The list shows the institutional affiliation of each committee member at the time of service.

Test Development Committee

J. Duane Hoover, Texas Tech University, TX
Chair
 Dorothy N. Harlow University of South Florida, FL

Joseph L. Massie University of Kentucky, KY
 David R. McKay Monroe County Community College, MI
 Stanley J. Seimer Syracuse University, NY

Working with a test specialist at Educational Testing Service (ETS) during 1978-79, this committee determined the content specifications of the new forms, and then prepared, selected, and reviewed carefully all items on each form of the test.

The primary objective of the Test Development Committee was to provide a test with good content validity. While the consensus of the committee members is that the test has high content validity for a typical introductory management course, the validity of the content for a specific course is best determined locally by carefully reviewing and comparing test content with the instructional content covered in the particular course.



The College Board

Description of Test

The Introduction to Management examination covers the material that is usually taught in a one-term course in the essentials of management and organization. The fact that such courses are offered by different types of institutions and in a number of fields other than business has been taken into account in the preparation of this examination. The test requires a knowledge of human resources and operational and functional aspects of management, but primary emphasis is placed on functional aspects of management.

The examination contains approximately 100 multiple-choice questions to be answered in two separately timed 45-minute sections.

Knowledge and Skills Required

Questions on the test require candidates to demonstrate one or more of the following abilities.

- Specific factual knowledge, recall, and general understanding of purposes, functions, and techniques of management (about 10 percent of the examination)
- Understanding of and ability to associate the meaning of specific terminology with important management ideas, processes, techniques, concepts, and elements (about 40 percent of the examination)
- Understanding of theory and significant underlying assumptions, concepts, and limitations of management data, including a comprehension of the rationale of procedures, methods, and analyses (about 40 percent of the examination)
- Application of knowledge, general concepts, and principles to specific problems (about 10 percent of the examination)

The subject matter of the Introduction to Management examination is drawn from the following topics.

Approximate Percent of Examination	
15%	Manpower and Human Resources Personnel administration Collective bargaining Human relations and motivation Manpower development
15%	Operational Aspects of Management Production planning and control PERT Quality controls Data processing and computer assistance Inventory control
60%	Functional Aspects of Management Planning Organizing Directing Controlling Authority Decision making Organization charts Leadership Structure Budgeting
10%	Miscellaneous Aspects of Management Historical aspects of management Social responsibilities of business Systems Contingency perspectives

Sample Test Questions

The following questions are provided to give an indication of the types of items that appear on the Introduction to Management examination. CLEP examinations

are designed so that average students completing a course in the subject can usually answer about half the questions correctly. Correct answers appear on page 5.

Directions: Each of the questions or incomplete statements below is followed by five suggested answers or completions. Select the one that is best in each case.

1. Which of the following words is NOT logically related to the others?
(A) Planning (B) Directing (C) Producing
(D) Controlling (E) Organizing
2. If a company increases its prices, the total revenue line on a break-even chart will
(A) shift upward over its entire length
(B) rise more steeply
(C) be unaffected
(D) rise less steeply
(E) shift downward over its entire length
3. Program Evaluation and Review Technique (PERT) is a system for
(A) developing the organization chart for a company
(B) scheduling and finding the critical path for production
(C) evaluating the performance of workers
(D) reviewing the overall financial condition of the company
(E) programming a computer
4. Which of the following is a correct statement about controlling as a management function?
(A) It can be performed independently of planning.
(B) It is performed only by the controller of an organization.
(C) It is more prevalent in business than in government.
(D) It assumes a certain approach to motivating employees.
(E) To work effectively, it must be closely related to planning.
5. Which of the following statements about the "exception principle" is (are) correct?
I. It states that those who take exception to a directive should report the basis of their disagreement or objection to their superior immediately.
II. It is concerned with correctly identifying what should be the logical "exceptions to every rule."
III. It is an important control concept.
(A) I only (B) II only (C) III only
(D) I and III only (E) II and III only
6. Decentralization tends to be encouraged by which of the following business trends?
I. Product diversification
II. Use of electronic computers
III. Geographical expansion of operations
(A) I only (B) II only (C) III only
(D) I and III only (E) II and III only
7. An effective program of work simplification would include all of the following steps EXCEPT:
(A) Begin the program by selecting a procedure wherein each unit of work is very time-consuming.
(B) Break the total procedure down into its basic elements.
(C) Question the elements for their essentiality.
(D) Eliminate, simplify, rearrange, and combine the elements so that they will fit the new procedure.
(E) Put the new method into practice with training and follow-up.
8. Which of the following can be best determined by consulting an organization chart?
(A) The size of the company
(B) The relationships of people
(C) The nature of work performed
(D) The relationship of positions
(E) The quality of management of the firm
9. Which of the following best illustrates informal organization?
(A) Line authority, such as that of the field marshal and battalion commander in the military
(B) Staff authority, such as that of personnel or cost control in manufacturing
(C) Functional authority, such as corporate supervision of the legal aspect of pension plans in branch plants
(D) The acceptance of authority by subordinates
(E) Groupings based on such things as technical ability, seniority, and personal influence
10. The number of subordinates who directly report to a superior refers to the manager's
(A) span of control (B) organizational role
(C) organizational structure (D) chain of command
(E) general staff

11. In the complex role of a staff person, the personnel administrator has LEAST need of
- (A) the capacity to appraise organizational performance against written statements of objectives and policies
 - (B) the ability to see the organization as a whole and each member in it as a total personality
 - (C) the ability to think clearly, practically, and with foresight
 - (D) the ability to command and to enforce rules and regulations among production employees
 - (E) skill in communication
12. The choice of organization structure to be used in a business should be
- (A) made by mutual agreement among all of the people affected
 - (B) based on consideration of the type of organization structures used by competitors
 - (C) subject to definite and fixed rules
 - (D) based on the objectives of each individual business
 - (E) made by organization specialists rather than managers
13. Behaviorists have charged traditional managers with a lack of concern for
- (A) formal organizational structure
 - (B) the quality of management decisions
 - (C) following organizational channels of communication
 - (D) the acceptability of management decisions to subordinates
 - (E) productivity through control
14. Which of the following typically requires more interpretation than the others?
- (A) A rule (B) A policy (C) A procedure
 - (D) They require equal interpretation.
 - (E) If formulated properly, none should require interpretation.
15. The concept of hierarchy of needs refers to which of the following?
- (A) Functional foremanship (B) Unity of command
 - (C) Line-staff conflict (D) Heuristic programming
 - (E) None of the above
16. Recruitment of college graduates is usually the responsibility of
- (A) first-line supervisors (B) divisional managers
 - (C) personnel managers (D) the placement office
 - (E) section chiefs
17. Which of the following statements best shows an understanding of international management?
- (A) Practically none of the managerial attitudes, techniques, and approaches found successful in the United States can be used in other (comparative management) societies.
 - (B) Although effective management plays an important role in the economic accomplishments of the United States, it has limited application in the developing countries of the world.
 - (C) An understanding of the cultural differences of various countries and their impact on the management process is essential to success in a multinational firm.
 - (D) Although some variations in cultures do exist, management fundamentals are universal and their application in different countries will generally be the same.
 - (E) An effective manager in the United States who has a basic understanding of human relations can easily be transferred to another country and perform successfully.
18. Douglas McGregor's "Theory Y" provides management with which of the following?
- (A) An understanding of those assumptions of human behavior deduced from the concept of the "rational economic man"
 - (B) A sound indication that production-centered supervisors are less effective than employee-centered supervisors
 - (C) The idea that employees are most effectively motivated when guided by a clearly expressed system of rewards and penalties
 - (D) A set of assumptions regarding human behavior implicit in much of the literature of traditional management
 - (E) None of the above
19. Henry Gantt, Frank Gilbreth, and Frederick Taylor are considered pioneers in the school of management generally referred to as the
- (A) management-process school
 - (B) empirical school
 - (C) scientific-management school
 - (D) behaviorist school
 - (E) social-system school
20. Preparation of which of the following is the most logical first step in developing an annual operating plan?
- (A) A sales forecast by product
 - (B) A production schedule by product
 - (C) A flow-of-funds statement by product
 - (D) A plant and equipment requirement forecast
 - (E) A pro forma income statement and balance sheet

21. A large span of control throughout an organization invariably results in
- (A) low morale
 - (B) high morale
 - (C) an excess work load for each manager
 - (D) a flat (horizontal) organizational structure
 - (E) a tall (vertical) organizational structure
22. Which of the following generally provide the impetus for the organization of informal work groups?
- (A) Contacts created by the work setting
 - (B) Individuals who oppose management
 - (C) Strong leaders who recruit members
 - (D) Individuals who know each other from nonwork activities
 - (E) Unions
23. Which of the following is an example of a line position in a manufacturing organization?
- (A) A sales manager concerned with selling a product in a given territory
 - (B) The head of research and development concerned with new products
 - (C) The controller concerned with establishing budgets
 - (D) The personnel manager concerned with employing workers
 - (E) The quality control manager concerned with maintaining quality standards in a production plant

24. Operations-research groups should consist of
- (A) mathematicians primarily
 - (B) statisticians primarily
 - (C) people trained in any quantitative discipline
 - (D) interdisciplinary teams
 - (E) middle and top management people

ANSWERS				
1. C	6. D	11. D	16. C	21. D
2. B	7. A	12. D	17. C	22. A
3. B	8. D	13. D	18. E	23. A
4. E	9. E	14. B	19. C	24. D
5. C	10. A	15. E	20. A	

Optional Free-Response Section

The optional free-response section of the Introduction to Management examination requires candidates to apply the principles of management to current issues and to bring together in a logical and coherent manner material drawn from different aspects of the subject. Candidates are asked to answer five questions within a 90-minute period, with local faculty assuming essay grading responsibility.

Score Interpretation Data

Table I below provides data to help interpret the scaled scores earned by candidates on the Introduction to Management test. The left-hand (unnumbered) column in Table I lists each possible test score in the middle range of the score scale, as well as selected scores in the upper and lower ranges. The scores range from 20 to 80 with a mean of 50. Raw scores are converted to scaled scores so that the scores for all test forms will be on the same scale and thus have the same value regardless of the form used. Because of substantial changes in the test specifications of the 1979 version, this examination was rescaled and renormed rather than equated to earlier forms.

The numbered column headings in Table I are explained in the following paragraphs.

Table I: Interpretive Data
CLEP Introduction to Management

Scaled Score	Percentile		Course Grade (3)	Range of Formula Scores (4)	Range of Correct Answers (5)
	(1)	(2)			
80	99				
75	99				
70	99				
65	94				
60	84			66-69	66-75
59	82				
58	79		A		
57	77				
56	74	(75)			
55	71	(70)		57-60	57-68
54	67				
53	64	(65)			
52	61	(60)	B		
51	57	(55)			
50	53			48-51	48-61
49	49	(50)			
48	46	(45)			
47	42	(40)	C		
46	38				
45	35	(35)		39-42	39-54
44	32				
43	30	(30)			
42	26	(25)			
41	23		D		
40	21			31-33	31-46
38	16		E		
35	9				
30	3				
25	1				
20	1				

Percentiles

- (1) The percentile for a particular scaled score is the percentage of students in the reference group that scored at or below that scaled score. The percentiles given in column (1) are based on the performance of 3,882 students completing introductory management courses in the spring of 1979 at the institutions listed on page 7.
- (2) Because not all possible numerical percentile ranks appear in column (1), the scaled scores corresponding to commonly used percentile ranks (every 5th from the 25th to the 75th) have been approximated by interpolation. Each of these percentiles is given in parentheses in column (2), across from its approximate scaled score. *Note that the percentile printed on the candidate's score report for a given scaled score will be the percentile given in column (1).*

Course Grades

- (3) To establish the concurrent validity of the Introduction to Management examination, the mean scaled scores of students in the 1979 reference group who earned final course grades of A, B, C, D, and E/F were computed. Of 2,475 students for whom course grades were available, 23% earned A, 35% B, 30% C, 9% D, and 3% E or F. Each course grade is indicated in column (3) opposite the mean scaled score of students who earned that grade.

As previously noted, the American Council on Education recommends that the minimum score for which credit is awarded should be the mean score of students in the reference group who earned a final course grade of C; this score (47) is printed in bold-face type in the table.

Range of Formula Scores

- (4) On all CLEP tests, a formula is applied to raw scores to correct for random guessing. This correction is designed so that someone who guessed at all test items would likely earn a formula score near zero. Formula scores are computed for this test by subtracting one-fourth of the number of wrong answers from the number of right answers. A range rather than a specific formula score is given in column (4) for selected scaled scores because the formula score corresponding to a particular scaled score can vary from form to form and, even on a

single form, two or more different formula scores sometimes convert to the same scaled score.

Range of Correct Answers

- (5) The number of correct answers given by candidates who earn each scaled score will fall within a given range, as illustrated in column (5). Because of the formula scoring process described in (4) above, and for the reasons given for formula scores falling in a range, the exact number of right answers given by a particular candidate cannot be determined from the scaled score. The candidate who gives the highest number of correct answers in the range will have answered incorrectly all the remaining questions on the test; one who gives the lowest number of correct answers in the range will have omitted all the remaining questions, answering none wrong. Both of these instances are unlikely; for a particular scaled score, most candidates will have given a number of right answers near the middle of the range.

The Reference Groups

The percentiles given in columns (1) and (2) of Table I are based on the performance of 3,882 students completing one-semester introductory management courses in the spring of 1979 at the institutions listed below. The validity (course grade) data reported in column (3) of Table I and the concurrent validity information on page 8 are based on the performance of 2,475 students from this administration for whom course grade data were available.

Anne Arundel Community College, MD
 Anderson College, SC
 Andrews University, MI
 Arkansas State University, AR
 Albright College, PA
 Augusta College, GA
 Briar Cliff College, IA
 Brookdale Community College, NJ
 Castleton State College, VT
 College of the Mainland, TX
 Dabney S. Lancaster Community College, VA
 Fairmount State College, WV
 George Mason University, VA
 Grand Rapids Junior College, MI
 Indiana State University-Evansville, IN
 James H. Madison University, VA
 Lansing Community College, MI
 The Loop College, IL
 Macomb County Community College, MI

Manatee Junior College, FL
 Mankato State University, MN
 Maple Woods Community College, MO
 Mississippi County Community College, AR
 Mississippi State University, MS
 Monroe County Community College, MI
 Montgomery College, MD
 Northeast Technical Community College, NE
 Presbyterian College, SC
 St. Peter's College, NJ
 St. Edward's University, TX
 St. Petersburg Junior College, FL
 State University of New York College at
 Plattsburgh, NY
 Sul Ross State University, TX
 Susquehanna University, PA
 Tarkio College, MO
 Tidewater Community College, VA
 Westchester Community College, NY
 Western State College, CO
 Tri-State University, IN
 University of Arkansas-Little Rock, AR
 University of Bridgeport, CT
 University of Texas-Arlington, TX
 University of Texas-El Paso, TX
 University of Science and Arts of Oklahoma, OK
 Tarrant County Junior College-South, TX
 Troy State University, AL
 Waubesa Community College, IL
 Western Connecticut State College, CT
 Westark Community College, AR
 Worcester State College, MA

The students in the reference group used for columns (1), (2), and (3) took forms of the Introduction to Management test that are currently in use. Because of substantial changes in the test specifications, the new forms could not be equated.

The data given in columns (4) and (5) of Table I apply to the same 1979 forms. Reliability, speededness, and item-difficulty data reported on pages 8 and 9 are derived from the spring 1979 administration of the current forms to college students.

Technical Information

Following is a brief discussion of the validity, reliability, and speededness of the Introduction to Management examination. A more detailed discussion of technical information and development procedures for CLEP examinations is available from the address given on page 1.

Validity

Test validity is established by determining whether a particular test measures what it is supposed to measure. One type of validity, concurrent validity, is addressed by determining whether the students who earn better grades in a course also score higher in the appropriate CLEP examination after taking the course. Correlations between final course grades and CLEP scores can be expected to fall in the middle range, i.e., between .30 and .70. For current forms of Introduction to Management, this correlation is .56.

In addition to the correlation between test scores and course grades, the mean score of students earning each course grade is computed. These are the mean scores appearing opposite the course grades in column (3) of Table I.

The content validity of the Introduction to Management examination was addressed by the Test Development Committee (see page i), through its efforts to develop a test that reflected the content of introductory management courses at most colleges. Content validity should be further addressed by each institution considering use of the examination. Appropriate faculty members should review the content outline and sample questions to ensure that the test covers, to a reasonable degree, the core material taught in their own courses.

Reliability

Test reliability is the degree to which a particular test consistently measures what it is intended to measure. The reliability coefficient is an estimate of the proportion of the variance in candidate scores that is due to true differences in ability rather than fluctuations due to chance or factors other than those being tested. Reliability estimates for CLEP examinations are computed by Kuder-Richardson Formula 20 (KR-20), adapted for use with formula scores, and these estimates are expected to be near .90. The reliability estimates for the two current forms of Introduction to Management are both .91.

The standard error of measurement (SEM) is another measure of test reliability. It is expressed in the same units as those of the score scale of the test, and indicates the probable range of discrepancy between a candidate's actual score and true score (i.e., the score that would be earned if the test could measure with perfect accuracy). There is a 67 percent chance that the candidate's actual score is within one SEM of his or her true score, and a 95 percent chance that it is within two SEMs of his or her true score. The SEMs for the two current forms of the Introduction to Management examination are 2.96 and 3.00.

Speededness

A test is speeded to the extent that performance on it is determined by the number of questions candidates have time to answer. Within a separately timed section of a test, all questions beyond the last question answered by a candidate are classified as not reached. A test may be regarded as essentially unspeded if at least 80 percent of the group reach the last question and if virtually everyone reaches three-quarters of the questions. By these criteria, from the data given in Table II below, there appears to be no evidence of speededness in the Introduction to Management test.

The ratio of the not-reached (NR) variance to the total score variance also yields a measure of speededness as a factor in determining scores. A test can be considered unspeded if the ratio of the NR variance to the score variance is less than .25. The low ratios between the NR and total score variances shown in Table II indicate that speed is not a major factor in determining scores on either form of this test.

Table II: Speededness of Sections*
CLEP Introduction to Management

	Form 1		Form 2	
	Sec. I	Sec. II	Sec. I	Sec. II
Percent completing section	84.5	87.1	86.2	74.4
Percent completing 75% of section	99.7	99.7	99.8	99.3
Ratio of not-reached variance to score variance	.02	.03	.03	.04
Number of items reached by 80% of the candidates	50	50	50	49
Total number of questions	50	50	50	50

*Speededness data are given for two test sections because the test is administered in two separately timed sections.

Item Characteristics

An index used by Educational Testing Service to describe item difficulty is called "delta"; it is a statistical transformation of the percent answering the item correctly. Deltas range from 6.0 (a very easy item) to 20.0 (a very difficult item).

The mean observed (or raw) difficulty levels or deltas for both forms of the test are above the middle-difficulty delta of 11.9 established for tests with five-choice items. Mean observed deltas of 12.0 and 12.2 for forms 1 and 2, respectively, indicate that the tests were moder-

ately difficult for the 3,882 students who served as the reference population at the spring 1979 administration. See Table III, below, for item-difficulty distributions.

Another item-analysis index used at ETS is the biserial coefficient of correlation between an item and an appropriate criterion, usually the total score on the test. The biserial correlation is clearly affected by the overall content of the test because the criterion is nothing more than the sum of the scores on all items that constitute the total. Distributions of biserial correlations of item scores with corresponding section and total scores are presented in Table IV on the right.

Table III: Distributions of Items by Difficulty Level (Delta) CLEP Introduction to Management

Observed Delta	Form 1			Form 2		
	Sec. I	Sec. II	Total	Sec. I	Sec. II	Total
19.0 up						
18.0-18.9			1		1	1
17.0-17.9	1		2		1	1
16.0-16.9	2		5		2	4
15.0-15.9	1	4	5	2	2	4
14.0-14.9	5	7	12	4	6	10
13.0-13.9	7	7	14	12	11	23
12.0-12.9	9	7	16	11	7	18
11.0-11.9	9	9	18	8	13	21
10.0-10.9	6	9	15	3	4	7
9.0- 9.9	5	4	9	4	3	7
8.0- 8.9	3	1	4	4		4
7.0- 7.9	2		2	1	2	3
6.9 down		2	2	1		1
Total	50	50	100	50	50	100
Mean	11.9	12.1	12.0	12.0	12.4	12.2
Standard Deviation	2.2	2.1	2.2	2.0	2.0	2.1

Table IV: Distributions of Items by Biserial Correlation of Item Score with Section and Total Scores CLEP Introduction to Management

r _{bis}	Form 1			Form 2		
	Sec. I	Sec. II	Total	Sec. I	Sec. II	Total
.90-.99						
.80-.89			1			
.70-.79		1	3	1		1
.60-.69	3		3			
.50-.59	7	16	23	7	14	21
.40-.49	19	19	38	20	19	39
.30-.39	18	8	26	16	13	29
.20-.29	3	5	8	5	4	9
.10-.19		1	1			
.00-.09						
Negative						
Total	50	50	100	49	50	99
Not Computed*				1		1
Mean	.42	.45	.43	.41	.43	.42
Standard Deviation	.09	.11	.10	.09	.09	.09

*Biserial correlations are not computed for questions that are reached by less than half of the group or that are answered correctly by more than 95 percent or less than 5 percent of the group.

Test Summary

The Technical Data Summary, Table V on page 10, presents comparative data for the current forms of the Subject Examination in Introduction to Management. Total-group, item, and test statistics are provided.

Table V: Technical Data Summary
CLEP Introduction to Management

Form Administration Number in Sample	Form 1 Spring 1979 1,950*	Form 2 Spring 1979 1,925*
TOTAL-GROUP STATISTICS		
Formala-Score Information		
Mean	50.54	48.66
Standard Deviation (S.D.)	17.63	17.65
Possible Range	-25-100	-25-100
Obtained Range	-7.96	-6.95
Median	51.36	49.18
Scaled-Score Information		
Mean	50	50
Standard Deviation (S.D.)	10	10
Possible Range	20-80	20-80
Obtained Range	20-76	20-76
Median	50.47	50.30
ITEM STATISTICS		
Number of Items		
Section I	50	50
Section II	50	50
Total Test	100	100
Mean Observed Delta	12.0	12.2
S.D. Observed Delta	2.2	2.1
Mean R-Biserial		
Section I	.42	.41
Section II	.45	.43
Total Test	.43	.42
S.D. R-Biserial		
Section I	.09	.09
Section II	.11	.09
Total Test	.10	.09
No. R-Biserial < .20		
Section I	0	0
Section II	1	0
Total Test	1	0
TEST STATISTICS		
Reliability	.913	.910
Raw Std. Error of Measurement	5.2	5.2
Scaled Std. Error of Measurement	2.96	3.00
SPECIAL SCORE DATA		
Mean Rights		
Section I	29.62	29.31
Section II	28.93	27.54
Total Test	58.55	56.85
Mean Wrongs		
Section I	16.21	15.83
Section II	16.30	17.44
Total Test	32.51	33.27
Mean Omits		
Section I	3.85	4.44
Section II	4.37	4.43
Total Test	8.22	8.89
% in Chance-Score Range		
Section I	2	3
Section II	4	4
Mean Percent Correct		
Section I	59	59
Section II	58	55
Total Test	59	57
Speededness (Section I)		
% Completing Section	84.5	86.2
% Completing 75% of Section	99.7	99.8
No. of Items Reached by 80%	50	50
No. of Items in Section	50	50
Mean Not-Reached	.32	.41
S.D. Not-Reached	1.32	1.46
NR Variance/Score Variance	.02	.03
Speededness (Section II)		
% Completing Section	87.1	74.4
% Completing 75% of Section	99.7	99.3
No. of Items Reached by 80%	50	49
No. of Items in Section	50	50
Mean Not-Reached	.41	.56
S.D. Not-Reached	1.64	1.89
NR Variance/Score Variance	.03	.04

*This actual number of cases used for analysis is slightly less than the total sample, due to rounding and low cases that did not meet established statistical requirements.

APPENDIX B

SAMPLE STUDENT EVALUATION OF INSTRUCTION FORM

APPENDIX C

SAMPLE OF MEMO TO STUDENTS AT TIME OF PRETEST

To: All Students Enrolled in Principles of
Management

From: Jim Jones, Associate Professor

Date: August 6, 1988

Subject: Research Project

There is a research project being conducted in the Business Administration Department this semester in which you will be involved. You will be taking two tests and evaluating your professor.

The first test is the one you are taking now. It is to get some idea of the level of management knowledge you have coming into the course. This test will not be considered as part of your grade for the course, but you are asked to do your best so that the results are as true as possible.

The other test you will take will be very similar to this one. It will be taken at the time of the final exam for the course. It will count as part, or all, of your final exam grade. How much it will count will be up to your professor.

The evaluation you do of your professor will be toward the end of the semester. Great care is taken by the college to make sure your personal rating is not made known to the professor. A special technique using a code number will be used so that your test scores can be matched with the rating for purposes of the research. Only the Department Chairperson or the Project Director will have access to this information.

If you have any questions, please contact the project director, Associate Professor Jim Jones, at 886-1333.

APPENDIX D

SAMPLE POSTTEST MEMO TO STUDENTS AND DATA SHEET FOR
DEMOGRAPHIC INFORMATION

To: All Students Completing Principles of Management
 Subject: CLEP final exam

You completed a test at the beginning of the semester to determine the level of your knowledge of the Principles of Management. The first test you took was not to be considered as part of your grade. The test you are about to take is all, or part, of the final exam for the course. This final exam does count toward your grade.

Before starting the test, please complete the information requested below. The information will be treated as confidential and will only be used for the research project.

NAME (PLEASE PRINT) _____

SOCIAL SECURITY NUMBER _____

AGE: _____ SEX: MALE _____ FEMALE _____ MAJOR: _____

ETHNIC ORIGIN:
 _____ AMERICAN INDIAN _____ ASIAN _____ HISPANIC
 _____ ANGLO _____ BLACK _____ OTHER _____

CHECK THE CLASS SECTION YOU ARE IN:

SEC	TIME	DAYS	LOCATE	Professor
_____01	08:00 AM	MWF	PB401	Nickeson
_____02	09:00 AM	MWF	PB601	Sifrit
_____03	10:00 AM	MWF	BA213	Combs
_____04	11:00 AM	MWF	PB401	Nickeson
_____05	08:00 AM	TT	PB601	Jones
_____06	09:30 AM	TT	PB602	Simmons
_____50	06:40 PM	MW	BA220	Guion
_____51	05:15 PM	TT	BA215	Tines
_____71	Telecourse			Sifrit
_____84	03:45 PM	TT	BLDG10	Blomquist

Turn in this page with the finished test.

SELECTED BIBLIOGRAPHY

Books

- The Random House College Dictionary, Revised Edition. New York, N.Y.: Random House, Inc.. 1980.
- Bernardin, John, and Richard W. Beatty. Performance Appraisal: Assessing Human Behavior at Work. Boston: Kent Publishing Co.. 1984.
- Borg, Walter R., and Meredith D. Gall. Educational Research, An Introduction, 4th Edition. New York: Longman, Inc.. 1983.
- Boyd, James E., E. F. Schietinger. Faculty Evaluation Procedures in Southern Colleges and Universities. Atlanta: Southern Regional Education Board. 1976.
- Carroll, Stephen J., and Craig E. Schneier. Performance Appraisal and Review Systems; The Identification, Measurement, and Development of Performance in Organizations. Glenview, Il.: Scott, Foresman and Company. 1982.
- Cook, David L. "An Objective Component for the Evaluation of Teaching". Ed.D diss., University of Tennessee. 1985.
- Cummings, L.L. and Donald P. Schwab. Performance in Organizations. Glenview, Il.: Scott, Foresman & Co.. 1973.
- Fournier, F., and Associates. Performance Appraisal Design Manual. Bridgewater, N.J.: Fournier, F. & Associates. 1983.
- Guba, Egon G., and Yvonna Lincoln. Effective Evaluation. San Francisco, Ca.: Jossey-Bass Publishing. 1981.
- Huck, Schuyler W., William H. Cormier and William G. Bounds, Jr.. Reading Statistics and Research. New York: Harper and Row. 1974.
- Kellogg, Marion S., What To Do About Performance Appraisal. New York, N.Y.: American Management Association. 1965.
- Latham, Gary P., and Kenneth N. Wexley. Increasing Productivity Through Performance Appraisal. Reading, Mass.: Addison-Wesley Publishing Co.. 1981.
- Miller, Richard I. Evaluating Faculty Performance. San Francisco: Jossey-Bass, Inc.. 1972.

- Miller, Richard I., Developing Programs for Faculty Evaluation. San Francisco: Jossey-Bass Publishers. 1974.
- Mondy, R. Wayne, and Robert M. Noe III. Personnel: The Management of Human Resources, Second Edition. Boston: Allyn and Bacon, Inc.. 1984.
- Norusis, Marija J., The SPSS Guide to Data Analysis for SPSSx. Chicago: SPSS Inc.. 1988.
- Patten Jr., Thomas H. A Manager's Guide to Performance Appraisal: Pride, Prejudice, and the Law of Equal Opportunity. The Free Press. 1982: New York. .
- Pedhazur, Elazar J., Multiple Regression in Behavioral Research; Explanation and Prediction, Second Edition. New York: Holt, Rinehart and Winston. 1982.
- Price, Marianne. "The Development of Behaviorally Anchored Rating Scales for the Performance Evaluation of Specific Education Teachers.". Ed.D. Dissertation, Temple University. (1979)
- Sedlin, Peter., Changing Practices in Faculty Evaluation. San Francisco: Jossey-Bass Publishers. 1984.
- Smith, Al. Lawrence H. Poole, Donald Dellow, Richard Behrendt, and William E. Cashin. Evaluating Faculty and Staff: New Directions for Community College Series #41. San Francisco: Jossey-Bass. 1983.

Journals

- Atkin, Robert S.. and Edward J. Conlon. "Behaviorally Anchored Rating Scales: Some Theoretical Issues". Academy of Management Review. vol 3 (January 1978): pp. 119-128
- Beer, Michael. "Performance Appraisal: Dilemmas and Possibilities". Organizational Dynamics. 1981 (Winter 1981): pp 24-36
- Bernardin, H. John. and Patricia Cain Smith. "A clarification of some issues regarding the development and use of behaviorally anchored rating scales (BARS)". Journal of Applied Psychology. 66(4) (1981): pp. 458-463
- Cascio, Wayne F.. and H. John Bernardin. "Implications of Performance Appraisal Litigation for Personnel

- Decisions". Personnel Psychology. Vol 34 No 2 (Summer 1981): pp 221-226
- Cohen, P. A.. . "Student Ratings of Instruction and Student Achievement: A Meta-analysis of Multisection Validity Studies". Review of Educational Research. Vol 51 (1981): 281-309
- Coker, Homer. , Donald M. Medley and Robert Soar. "How Valid Are Expert Opinions About Effective Teaching?". Phi Delta Kappan. October, 1980. pp. 131-135.
- Giglioni, G.B.. J.B. Giglioni and J.A. Bryant. "Performance Appraisal: Here Comes the Judge". California Management Review. Vol XXIV, No. 2 (Winter 1981): pp. 14-23
- Green, Samuel B.. and others. "Shortcut Methods for Deriving Behaviorally Anchored Rating Scales". Educational and Psychological Measurement. No.41 (1981): pp. 761-775
- Harari, Oren. and Sheldon Zedeck. "Development of Behaviorally Anchored Scales for the Evaluation of Faculty Teaching". Journal of Applied Psychology. Vol. 58, No. 2 (1973): p261-265
- Hom, Peter W.. and others. "Effectiveness of performance feedback from behaviorally anchored rating scales.". Journal of Applied Psychology. Vol 67(5) (Oct, 1982): pp. 568-576
- Kaufman, Brian J.. "Departmental Differences in Student Perceptions of 'Ideal' Teaching". Paper presented at the Annual Meeting of the Southeastern Psychological Association. March 26, 1981.
- Kaufman, Brian J.. and Joseph Madden. "The Development of Behaviorally Anchored Rating Scales for Student Evaluation of College Teaching Effectiveness. ". Paper presented at the 51st Annual Meeting of the Eastern Psychological Association. April 9-12, 1980.
- Kearney, William J.. "Behaviorally Anchored Rating Scales--MBO's Missing Ingredient.". Personnel Journal. v58 n1 (Jan. 1979): p20-25
- King, L.M., J.E. Hunter and F. L. Schmidt. "Halo in a multidimensional forced-choice performance evaluation scale.". Journal of Applied Psychology. Vol 65 pp. 507-516
- Kleiman, Lawrence S.. and Richard L. Durham. "Performance Appraisal, Promotion, and the Courts: A Critical

- Review". Personnel Psychology. Vol 34 No. 1 (Spring 1981): pp. 103-121
- Laird, A.. and P. C. Clampitt. "Effective Performance Appraisal; Viewpoints From Managers". Journal of Business Communications. Vol 22 (Summer 1985): pp. 49-57
- Landy, Frank J.. and Janet Barnes. "Scaling Behavioral Anchors.". Applied Psychological Measurement., v3 n2 (Spr): 979
- Rarick, C.A.. and G. Baxter. "Behaviorally Anchored Rating Scales (BARS): an effective performance appraisal approach". Advanced Management Journal. Vol 51:36-9 (Winter 1986):
- Reinhardt, C.. . "The State of Performance Appraisal: A Literature Review". Human Resource Planning. Vol 8 No. 2:104 (October 1985)
- Schneier, Craig Eric. and Richard W.Beatty. "Developing Behaviorally-Anchored Rating Scales (BARS).". Personnel Administrator. v24 n8 (Aug 1979): p59-68
- Schneier, Craig Eric. and Richard Beatty. "Combining BARS and MBO: Using an Appraisal System to Diagnose Performance Problems.". Personnel Administrator., v24 n9 (Sep, 1979): p51-60
- Schwab, Donald P., Herbert G. Heneman III and Thomas A.DeCotiis. "Behaviorally Anchored Rating Scales: A Review of the Literature". Personnel Psychology. Vol 28, 4 (Winter, 1975): pp. 549-62
- Schwind, Hermann F., "New Ways to Evaluate Teaching and Training Effectiveness. ". Paper presented at the Adult Education Research Conference. (1977)
- Schwind, Hermann F., "A New Behavior Oriented Evaluation Scale to Assess Teaching and Training Effectiveness. ". Paper presented at the Evaluation Research Society Meeting.. (1977)
- Siegel, L., "A data-based scheme for evaluating faculty performance.". Research in Higher Education. vol 8 (1978): p255-271
- Silverman, Stanley B.. and Kenneth Wexley. "Reaction of employees to performance appraisal interviews as a function of their participation in rating scale

development.". Personnel Psychology. Vol 37(4) (1984): pp. 703-710

Smith, P.C.. and L.M. Kendall. "Retranslation of Expectations: An approach to the construction of unambiguous anchors for rating scales.". Journal of Applied Psychology. Vol 47 (1963): pp. 149-155

Soar, Robert S., Donald M. Medley and Homer Coker. "Teacher evaluation: a critique of currently used methods.". Phi Delta Kappan. (Dec. 1983). p. 239(13)

Taylor, R.L.. and R.A. Zawacki. "Trends in Performance Appraisal: Guidelines for Managers". Personnel Administrator. Vol. 29 No. 3 (Mar. 1984): pp. 71-80

Wilson, T.L., H.R. Dodge and D.H. Mathews. "Value Added Testing Is Likely To Expand". Marketing News. (July 3, 1987). p. 25