

INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.
2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.
3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.
4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.
5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms

300 North Zeeb Road
Ann Arbor, Michigan 48106

74-20,257

ANNADALE, Arthur David, 1942-
THE DEVELOPMENT AND VALIDATION OF AN INSTRUMENT
TO MEASURE TEACHING EFFECTIVENESS.

The University of Arizona, Ph.D., 1974
Education, psychology

University Microfilms, A XEROX Company, Ann Arbor, Michigan

© 1974

ARTHUR DAVID ANNADALE

ALL RIGHTS RESERVED

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED.

THE DEVELOPMENT AND VALIDATION OF AN INSTRUMENT
TO MEASURE TEACHING EFFECTIVENESS

by

Arthur David Annadale

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF EDUCATIONAL PSYCHOLOGY
In Partial Fulfillment of the Requirements
For the Degree of
DOCTOR OF PHILOSOPHY
In the Graduate College
THE UNIVERSITY OF ARIZONA

1 9 7 4

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

I hereby recommend that this dissertation prepared under my
direction by Arthur David Annadale
entitled The Development and Validation of an Instru-
ment to Measure Teaching Effectiveness
be accepted as fulfilling the dissertation requirement of the
degree of Doctor of Philosophy

Shitala P. Mishra 3/25/74
Dissertation Director Date

After inspection of the final copy of the dissertation, the
following members of the Final Examination Committee concur in
its approval and recommend its acceptance:*

<u>Robert C. Adams</u>	<u>3/25/74</u>
<u>Ruth W. Kupperley</u>	<u>3/25/74</u>
<u>Richard L. Brinkman</u>	<u>3/25/74</u>
_____	_____
_____	_____

*This approval and acceptance is contingent on the candidate's
adequate performance and defense of this dissertation at the
final oral examination. The inclusion of this sheet bound into
the library copy of the dissertation is evidence of satisfactory
performance at the final examination.

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: Arthur David Annadali

ACKNOWLEDGMENTS

For their abiding belief in my potential I especially want to thank Drs. Sigmund Hsiao, Shitala P. Mishra, and Dean F. Robert Paulsen. Without their timely assistance and faithful encouragement my doctoral program may have never been completed.

I owe much to my family and friends for their patient understanding. Each of them knows their unique contribution.

I want to thank the enthusiastic members of the Teacher Evaluation Committee of the College of Education for their investment of time, energy, and cooperation. Also, a warm note of gratitude is extended to Drs. Ruth W. Kingsley, Robert E. Calmes, Richard L. Erickson, Darrell L. Sabers, and Paul M. Allen for their valuable suggestions and help. Special thanks is given to Drs. Don Hoyt, Wilbert McKeachie, and James Popham for their challenging recommendations.

To Drs. Carl Rogers and Abraham Maslow whose writings have been a source of continual inspiration throughout my career and personal development, I want to extend a heartfelt thanks. Maslow once said that great books are never written unless someone attempts them and dreams. In this respect, for allowing me to dream in pursuing this dissertation topic, I want to thank Dr. Mishra. His creative ingenuity, research competence, and humble wisdom have made a dynamic impact on my life as well as on this dissertation.

To all of these people acknowledged I owe so much more than words can convey.

TABLE OF CONTENTS

		Page
	LIST OF TABLES.	viii
	ABSTRACT.	x
1	INTRODUCTION.	1
	Importance of the Problem	2
	Purpose of the Study.	4
	Rationale for the Study	5
	Definition of Terms	9
	Assumptions	9
	Summary	9
2	REVIEW OF RELATED LITERATURE.	11
	Historical Development of Rating Scales	11
	Informal Student Evaluations.	11
	Informal Administrative Evaluations	12
	Formal Student Evaluations.	13
	Empirical Perspective	14
	Research Related to the Use of Student Evaluation of Instruction	15
	Evaluating Teaching	17
	Research Related to the Psychometric Qualities of Rating Instruments	20
	Reliability	20
	Validity.	21
	The Content of Rating Scales Used by Previous Researchers.	22
	Teacher Personality Characteristics	22
	Teaching Skill.	24
	Organizational Structure of the Course.	25
	Teacher-Student Interactions.	26
	Research Findings and Theoretical Positions Regarding Effective Teaching.	26
	Defining Effective Teaching	26
	A Possible Criterion of Teaching Effec- tiveness.	29
	A Tentative Definition of Effective Teachers.	30
	Summary	32

TABLE OF CONTENTS--Continued

	Page	
3	METHODOLOGY	33
	Samples	33
	Teachers.	33
	Students.	34
	Development of Instrument	35
	Formation of Committee on the	
	Evaluation of Teaching	35
	Selection of Items for the First	
	Draft of the Instrument.	36
	Description of the Initial Instrument	38
	Administering Initial Instrument.	42
	Selection of Items for the Final	
	Draft of the Instrument.	42
	Administration of Instrument and Data	
	Collection.	43
	Administration of the 52-Item	
	Instrument	43
	Description of the Final Instrument	44
	Administration of the Developed	
	Instrument	46
	A Scheme of Evaluation Based on	
	Teachers' Preferences.	46
	Scheme for Data Analysis.	47
	Analysis for Determining Psycho-	
	metric Qualities of the Instrument	47
	Hypotheses to Be Tested	48
	Summary	49
4	RESULTS	50
	Results Related to Item Characteristics	50
	Validity Data	53
	Relationship between Perceived	
	Learning and Student Ratings	
	of their Instructors	57
	Reliability Data.	58
	Reliability of the 47-Item Instrument	58
	Intercorrelations of the 47-Item Form	58
	College Norms with Final Form of Instrument	62
	Results Related to Hypotheses	62
	Summary	66
5	SUMMARY, CONCLUSIONS, IMPLICATIONS, LIMITATIONS, AND RECOMMENDATIONS	67

TABLE OF CONTENTS--Continued

	Page
Summary	67
Conclusions	68
Implications.	70
Limitations	78
Recommendations	79
APPENDIX A: 69-ITEM INITIAL INSTRUMENT.	82
APPENDIX B: 52-ITEM INTERMEDIATE INSTRUMENT	89
APPENDIX C: 47-ITEM FINAL FORM.	94
APPENDIX D: LETTER TO FACULTY REQUESTING VOLUNTEERS FOR THE TEACHER EVALUATION.	100
APPENDIX E: LETTER TO FACULTY VOLUNTEERS FOR ADMINISTERING THE 47-ITEM INSTRU- MENT.	104
APPENDIX F: LETTER TO FACULTY VOLUNTEERS ASKING FOR THE SELECTION OF RELEVANT ITEMS	107
REFERENCES.	109

LIST OF TABLES

Table		Page
1	Departmental Affiliation of Sampled Teachers	34
2	Number and Sizes of Sampled Classes.	35
3	Teacher Evaluation Committee Members' Departmental Affiliations and Competencies	36
4	Items Comprising Seven Hypothetical Factors of the 69-Item Instrument.	40
5	Items Comprising Seven Hypothetical Factors of the 47-Item Instrument.	45
6	Means and Standard Deviations of 52 Items of the 69-Item Initial Instrument.	51
7	T-Ratios Obtained from Comparing Faculty and Student Perceptions of Appropriateness of the Items on the 52-Item Intermediate Instrument.	55
8	Items Not Meeting the 5.00 Criterion Level for Both Faculty and Student Groups on the 52-Item Instrument.	56
9	T Test of Frequency of Responses for Item 40 (1 or 2) in Relationship to Items 9-35 Combined	58
10	Intercorrelations of Items 9-13 of the 47-Item Instrument	59
11	Items 9-13 Means Correlated with Means of Items 9-35 Combined.	60
12	Intercorrelations of Item Means 41-47 with Ratings of Items 9-35 Combined	61
13	Item Means and Standard Deviations of the 47-Item Form	63

LIST OF TABLES--Continued

Table		Page
14	ANOVA Summary Related to Class Size and Faculty Selection of Items	64
15	Mean Ratings on Faculty Selected and Non- selected Items in Relationship to Class Size	65

ABSTRACT

The major purpose of this study was to construct an evaluative instrument for measuring teaching effectiveness. More specifically, an attempt was made to develop a reasonably reliable and valid rating device which could be used at the college level. Such an instrument might be used for improving teaching and for providing an objective source of data in making administrative decisions concerning teachers.

The review of the literature indicated that historically evaluations of teachers have been made on a subjective, unsystematic basis. In the past, nonobjective, informal student and administrative judgments have been the most common procedures used to assess teaching effectiveness. Quantity of research productivity in contrast to teacher productivity has been another unsystematic criterion used to gauge the quality of teaching at the university or college level. Research studies demonstrated that teaching activities may be improved through systematic feedback from student ratings. The results of the literature review further indicated that student ratings of teaching have high reliability and acceptable validity and that quality of teaching can be assessed by student ratings.

Certain teaching activities and qualities were assumed to be important dimensions contributing to effective teaching. The 47-item instrument developed and validated in this study is based on 227 items from five factor analytical investigations of teaching effectiveness.

During the fall semester, 1972, six faculty members of the Teacher Evaluation Committee of the College of Education at The University of Arizona helped determine the content validity of the instrument. Each committee member selected from the 227-item domain only those items believed to be related to effective teaching. With the approval of this committee a 69-item instrument was constructed for a pilot study. Using these 69 items, 1900 students rated 50 volunteer teachers. These ratings provided data for developing a final instrument to measure teaching effectiveness.

From the main body of the initial 69 items the first 52 items, which depicted effective teaching, were resubmitted to a representative group of faculty (N=86) and students (N=169) of the College of Education. In general, items rated mutually high (rating range of 5.00 - 6.00) by both faculty and students on this 52-item intermediate instrument were considered for retention in the final form. On this basis the criterion validity of the instrument was determined.

During the spring semester, 1973, for the second phase of the instrument's development 50 volunteer faculty and 1650 of their students from the College of Education helped determine the psychometric properties of the final form. On this 47-item final form, the reliability of its 27 items depicting effective teaching activities and qualities was found to be .85. Other item correlations were computed, demonstrating high internal consistency of the instrument.

In using the final form of the instrument two hypotheses were tested. The first hypothesis stated that there would be no significant

difference in student ratings of instructors among small, medium, and large classes. This hypothesis was found to be tenable. Thus, no significant differences were found among the small, medium, and large classes in relationship to student ratings of teaching effectiveness. The second hypothesis stated that there would be no significant difference in student ratings of instructors on faculty selected versus non-selected items on the rating forms. The instructors, in general, obtained significantly higher ratings on self-selected items than on nonselected items. Therefore, the second hypothesis was rejected.

The results from this study suggest that a fair, systematic evaluation procedure might be enhanced considerably by the use of rating items which are consistent with teachers' instructional objectives and philosophies. The data of this study indicated that the developed instrument has high reliability and adequate validity for measuring teaching effectiveness. Such an instrument as this one might be used as a feedback source for improving instruction and/or as an administrative aid for making objective decisions concerning teacher assignments, salaries, and promotions. However, the major value of such an evaluative device is believed to lie in its utility for instructor self-appraisal in the improvement of teaching.

CHAPTER 1

INTRODUCTION

In recent years strategies to facilitate learning in the classroom have attracted many researchers investigating teaching effectiveness. Research studies by Flanders (1960), Ryans (1960), Gage (1963), McKeachie (1963), Gagné (1965), and Popham (1971) are some examples of the efforts conducted to understand in depth the teaching-learning process.

As specific examples of these endeavors, Flanders (1960) and Ryans (1960) proposed that effective teachers can be identified by certain classroom procedures which enhance learning in students. Gage (1963) reviewed over a half century of research studies on teaching. A decade later Gage (1972) renewed his commitment to search for empirical dimensions of effective teaching. McKeachie (1963) and Gagné (1965) contributed a great amount of theoretical and empirical information to our understanding of teaching and learning.

It appears that the evaluation of teaching effectiveness is especially difficult in that learning can be accomplished in many different ways. An effective learning experience, therefore, is dependent upon a variety of factors. In particular, Gagné (1965) maintained that the outcome of learning is determined jointly by an individual instructor and a unique group of students working toward the accomplishment of educational objectives in a specific learning environment. A basic

responsibility of the teacher is to transmit subject content to the learners in such a way that their educational growth is maximized (Popham and Baker, 1970). Several instructional strategies have been suggested to accomplish such a task (Glaser, 1963). In addition, Gagné (1965) maintained that effective teaching and learning depend upon the curriculum, its content, the way in which it is structured, and the specific objectives which the teacher chooses for the course. Similarly, Bruner (1966) suggested that meaningful, conceptual learning is facilitated by effective teachers.

Importance of the Problem

In considering recent research contributions to our understanding of the educational process, it seems that one of the most potent variables often perceived as a manager and facilitator of learning experiences is the teacher. To a great extent, effective learning in children and school-age youngsters is influenced by competent teachers (Popham and Baker, 1970). At whatever age level we are concerned, the teacher is an important variable contributing to the efficacy of the educational process. However, there are various pertinent issues even today which cloud our understanding concerning teaching effectiveness.

It is believed that some instructional activities and procedures are more effective than others in facilitating student learning (Gage, 1963). Also, it has been proposed that if effective teaching were to become a more significant criterion for a teacher's academic advancement, teaching performance and the overall quality of instruction would improve (Eble, 1971; Hildebrand, Wilson, and Dienst (1971). Hoyt's

analysis (1969) of the problem pointed out why little progress has been made on this issue. Unless there is reasonable agreement on what basis (criteria) teaching effectiveness can be judged, it will be difficult to develop dependable guidelines for improving instruction and for facilitating educational growth in students.

There are various ways to determine teaching effectiveness. The simplest and quickest way to measure teaching effectiveness is to use student ratings of teachers' classroom performance. Costin, Greenough, and Menges (1971) suggested that the results obtained from student ratings can be utilized meaningfully in providing comprehensive feedback to the instructors which they may not be able to receive from students even on a one-to-one personal basis. Also, student evaluations of teacher competencies could provide faculty members with evidence of their teaching ability for administrators who make decisions concerning promotions, salary increases, and teaching assignments. Information could be supplied to the department and college on areas of strengths and weaknesses in the curricula offerings. Thus, evaluative information, at least as far as student satisfaction is concerned, on the various new programs could be known. Furthermore, departmental and college-wide norms could be made available so that individual faculty ratings could be compared. Such systematic information might provide a valuable source to aid students in the selection of courses (Costin et al., 1971).

In addition to the possible uses of results obtained from student ratings, Hoyt (1969), Hartley and Hogan (1972), and Gage

(1972) advocated the use of student ratings as a valuable feedback procedure to improve instruction. Miller (1972) similarly suggested that sensitive evaluation techniques are needed to encourage good teaching. The teacher who is having difficulty and "not measuring up" must be identified so that he can be helped. Student ratings of teachers for use in improving the quality of instruction has been suggested by numerous researchers in recent years (Centra, 1973). It is believed that student evaluations are not less trustworthy than other available methods. A well designed student rating program, therefore, is very much needed to benefit the academic community.

Purpose of the Study

The major focus of this study was the development and validation of a student rating instrument that could be used for evaluating teacher effectiveness. The study involved a systematic attempt to develop a rating device to evaluate teaching effectiveness from the student's point of view within the contexts of instructional objectives and activities for the faculty in the College of Education at The University of Arizona. In order to attain this goal, areas of agreement in student-faculty perceptions (ratings) concerning what constitutes effective teaching were explored. Specific instructional activities perceived as being associated with effective teaching by both faculty and students served as a basis for the development of a rating instrument. A related, concomitant aim of the study was to characterize effective teaching in a reasonably reliable and valid manner consistent with the objectives of each instructor. This information might possibly be used by

instructors as one avenue of feedback to enhance their instructional competencies. Such data also might be used by administrators as a source of empirical evidence for making decisions concerning teacher promotions and salary increases.

Rationale for the Study

The use of a student rating form that has been carefully constructed and researched has several benefits. Teachers might receive a diversified, wide sample of evaluative feedback that is systematic, reasonably reliable and valid. A teacher evaluation program can be based on a solid research foundation and rationale. A comprehensive and systematic procedure for evaluating teaching effectiveness may possibly be used to carry out both formative and summative evaluations (Eble, 1971; Miller, 1972).

From various learning theories, models, or paradigms we learn that one of the best ways to maximize human potentialities is to receive knowledge of results of one's performance in relation to one's goals, objectives, or "target behaviors." In short, it seems wise to exercise some type of explicit, responsive feedback system (Wilhelms, 1967). The feedback-interaction model proposed by Runkel and advocated by Gage (1972) appears consistent with the research on effective teaching. This model may be used for instructional improvement.

Traditionally, teachers get some feedback about their pedagogical competencies based on class discussions, assignments, papers, and tests from students. Some instructors, however, do not solicit or

receive any kind of evaluative information from students. Such professors have eliminated a feedback channel and thus closed the door to one source of professional growth and change.

While much of the empirical data on teaching have contributed a wealth of information to our understanding of the teacher-learner process, in general, it fails to address itself adequately to the improvement of instruction. Empirical findings as well as theoretical positions support the view that education is a two-way process. If a teacher learns via student feedback the effects of his teaching activities and how he comes across to students, he most likely will become more effective in the classroom. Using feedback is one of the simplest yet most potent ways any organism learns and adapts to its environment. No one is a born teacher. A violinist has to practice and hear the result to improve. A teacher has to teach and receive feedback from students to improve his teaching skills.

Evaluation of faculty to improve instructional qualities has been the preeminent interest and guideline in the research efforts of Hoyt (1969) and others in recent years. Hoyt (1969), for example, believed that evaluation data are related intimately to professional development and self-renewal. A positive evaluation program can help uncover both strengths and weaknesses in faculty and help generate significant behavioral changes and learning increments in students (McNeil, 1971). An important fact in this regard is that the monetary investment in a tenured faculty member for a 25 year period represents an excess of \$600,000 (Miller, 1972). This sum of over a half million

dollars would seem to suggest the economic importance of systematically encouraging the improvement of each instructor on a continual, yearly basis.

To the largest segment of the higher education community--the students--teaching is usually perceived as the university's most important activity. It is not surprising, then, that student criticisms of higher education are often directed to the instructional function. Sanford (1968) and Scranton (1970) reported that a large number of students feel that the quality of their educational experience is impeded greatly by an unresponsive system. Sanford's Where Colleges Fail (1968, p. 36) elaborated the point that "we should encourage students to criticize the education we arrange for them, and that we should consider their suggestions thoughtfully. . .to listen to them and then to do what their actions show they need."

If one assumes that the purpose of education is to change student behavior as a result of some definite course of instruction, then an objective of educational research should be to find out what teaching procedures produce the desired behavioral changes in students. Once the desired behavioral changes can be defined and identified, the educational researcher can develop instruments to measure them and provide feedback for the benefit of both faculty and students. Students are constantly exposed to the variables associated with the teaching-learning process. They are, perhaps, the most logical evaluators of the quality and effectiveness of the course and instructor. Centra (1973) indicated that students are both the consumer and product of the

educational process and are able to evaluate their teachers more accurately than any other individual or group in the educational community. (As Ralph Nader has implied, the supporters of this approach advocated that one receives a better appraisal of the merits of a dinner from the consumers of the meal rather than the cook.) Furthermore, student ratings can indicate areas of rapport, degrees of communication, or the existence of problems that might not be discovered otherwise.

Today, the current interest in accountability prompts us to look for simple and empirical approaches to evaluation. Lessinger (1971) and Sciara and Jantz (1972) stated that the time and zeitgeist of our age are ripe for teacher accountability. They maintained that if educators do not assume professional responsibility for their teaching endeavors, someone else will. A dynamic accountability movement appears to be sweeping the country. Informed citizens are asking, "What are the students learning?" and "How good a job of education are our teachers doing?" State legislators and taxpayers are demanding an accounting of how well public monies are being spent. Miller (1972) contended that if evaluation is not done internally, it will be done externally by accountability experts working for legislators and governors. Colleges can be expected to become increasingly conscious of cost-effectiveness and possible budget reductions in the seventies (Lessinger, 1971; Miller, 1972).

The preceding rationale served as a guiding framework for this study in the development and validation of a student rating instrument to measure teaching effectiveness.

Definition of Terms

For the purpose of this study, teaching and teaching effectiveness are defined as follows:

1. Teaching is any interaction between a teacher and one or more students in which information is exchanged and learning is facilitated.
2. Teaching effectiveness refers to a value judgment to indicate the relative worth or value (some degree of goodness) in the quality of instruction as indicated by student perceptions (ratings) of their teachers as measured by the developed instrument in this study.

Assumptions

Two assumptions are important to this study: (1) since "effective teaching" involves a value judgment concerning the relative degree of goodness or badness of the quality of instruction, teachers need to be evaluated in the context of their unique styles of teaching, frames of references, and instructional objectives; and (2) teachers rated effective by students tend to facilitate more student learning in contrast to teachers rated ineffective.

Summary

This chapter contained an introduction to the problem, emphasizing the point that there is an acute need for developing an evaluative instrument for assessing teaching effectiveness. Two fundamental purposes of the study were stated to be: (1) the development and validation of a student rating scale to measure teaching effectiveness, and

(2) the use of obtained results for the professional growth of teachers to improve instruction. The chapter suggested the potential usefulness of a systematic teacher evaluation program. It was mentioned that teaching effectiveness can be assessed by the effects of teaching on student behavior as indicated by student perceptions (ratings). The chapter also presented the importance and rationale of the study. Finally, definitions of key terms of the study were given.

CHAPTER 2

REVIEW OF RELATED LITERATURE

This chapter will present a review of related research concerning the use of student ratings for evaluating teaching effectiveness. The organization of this chapter will be as follows: (1) historical background, (2) research related to the use of rating instruments, (3) research related to psychometric qualities of rating instruments, (4) the content of rating scales used by previous researchers, and (5) research findings and theoretical positions regarding effective teaching.

Historical Development of Rating Scales

Informal Student Evaluations

Over the years students have provided information concerning teaching effectiveness in both informal and formal ways. Metcalfe (1972) believed that teachers have always been evaluated by one way or another throughout civilization. As early as the fifteenth century at the University of Bologna students paid instructors based on their teaching abilities (Centra, 1973). In the United States as early as 1924, Harvard's "Confidential Guide" was distributed by students in their efforts to evaluate college teachers by the use of informal comments. Even today, unstructured verbal reports by students often become part of the hearsay evidence used to evaluate teachers. As an

example, at Wayne State and other well known universities, a small cadre of students sit in on classes for a week and write course evaluations in essay, narrative form (Eble, 1971). Flanagan's (1954) "critical incident" technique is another current, yet unsystematic, procedure used to evaluate teaching performance. Using this approach, students reported in essay form their reactions to key dramatic teaching episodes that occurred in a particular situation.

Informal Administrative Evaluations

In the past, the most frequently used method for evaluating an instructor's teaching was made by the second-hand judgment of a department chairman or dean (Gustad, 1961). This method of evaluating teachers has been criticized for being unsystematic, unreliable, and largely invalid (Hildebrand et al., 1971). By the same token, opinions solicited by professional colleagues, peers, or staff tend to be based on subjective evidence and do not constitute an adequate or representative sample.

Hildebrand (1972) illustrated the methodological problems, fallacious assumptions, and numerous pitfalls involved in the traditionally used administrative procedures to evaluate teaching. As one example, he suggested that a chairman or dean may be unqualified as a judge of effective classroom teaching. However, it is well known that classroom visitations are seldom made by administrators or even faculty colleagues in that such observations are resisted or resented by most teachers. Student information concerning an instructor's teaching

usually plays a minor part in deciding whether or not he is promoted (Eble, 1971). On this point Astin and Lee (1967, p. 299) commented, "If the ultimate measure of the teacher's effectiveness is his impact on the students--a view which few educators would dispute--it is unfortunate that those sources of information most likely to yield information about this influence are least likely used."

Over the years administrative opinions, faculty peer discussions, and research publications have outweighed quality of teaching as criteria for advancement on most college campuses (Hildebrand et al., 1971). Yet in a recent survey (Wilson, Gaff, and Bavry, 1970) of 1,000 faculty members at six diverse colleges and universities, 92 percent stated that teaching effectiveness should be "quite important" or "very important" as a criterion of advancement. No less than 72 percent of those respondents felt that their campuses should have a formal evaluation procedure.

Formal Student Evaluations

In this country formal attempts to evaluate teaching date back at least to 1915 (Metcalf, 1972). Some of the most thorough, systematic attempts to evaluate teachers appeared in the late 1920s at the University of Washington and Purdue University with the respective work of Guthrie and Remmers (Eble, 1971). Presently, as Gage (1972) and Centra (1973) indicated, a comprehensive, empirical approach using student ratings is the most reliable and valid means available for evaluating teaching. Current practice seems to favor a systematic rating form which covers the major theoretical and practical aspects of

effective teaching. In this way, such a questionnaire can be administered to large numbers of students in a relatively short time. The results can be printed on a computer printout with the ratings summarized on each individual item and normative data provided.

Empirical Perspective

Over the past 70 years numerous attempts have been made to identify, define, and measure what constitutes effective teaching. Ten years ago Gage (1963) in the Handbook of Research on Teaching reviewed a plethora of research findings on teaching effectiveness that spanned five decades. However, according to Biddle (1964), the results of these studies were quite confusing and disappointing. By 1963 hundreds of studies had been conducted, reporting thousands of correlation coefficients which were mainly nonsignificant and generally inconsistent from one source to the next. Mitzel (1960, p. 1481) reinforced this view by saying, "More than a half century of research effort has not yielded meaningful, measurable criteria around which the majority of the nation's educators can rally. No standards exist which are commonly agreed upon as the criteria of teacher effectiveness."

However, within the last decade a large body of empirical studies have yielded a renewed hope for identifying variables or processes related to effective teaching.

Gage (1972) mentioned that the earlier historical effort made great use of abstract, global ratings in contrast to the present day work that uses concrete, specific behaviors. The earlier work failed to delineate teacher roles, types of students, and educational

objectives. The more recent work recognized the need for clear specificity in these respects. Gage (1972) indicated that the earlier pessimism concerning a scientific understanding of teaching effectiveness is no longer warranted. With contemporary sophisticated research approaches, a scientific basis for understanding effective teaching and learning is becoming a solid, empirical reality.

Research Related to the Use of Student Evaluation of Instruction

In recent years systematic student evaluations of teachers and courses have been used widely at many colleges and universities. The results of these evaluations usually are seen only by instructors and are intended to help improve their teaching (Centra, 1972b; Gage, 1972). Underlying this intended use is the assumption that the instructors will use the information to modify and improve their teaching activities. As Centra (1972b, p. 1) stated, "It is an assumption that is open to question." However, Gage (1972) defended this assumption on the basis of what he called "equilibrium theory." Accordingly, when student feedback creates a condition of imbalance or dissonance in an instructor, one might expect the instructor to change in the direction desired by students in order to restore a condition of "balance" or "equilibrium." Following a reasonable lapse of time, such changes or improvements should be manifested in a second description or rating of teacher behavior. In fact, following the above reasoning, this type of student-teacher interaction and, in particular, teacher behavioral changes have been reported by several investigations. Hoyt (1969) and Centra (1972b)

presented evidence that student feedback via the form of student ratings has a positive effect on subsequent teaching performance. Tuckman and Oliver (1968) found that instructors who received feedback showed greater gains in student ratings after a 12-week interval than those instructors who received no feedback. Similar evidence was reported by Bryan (1963) and Gage, Runkel, and Chatterjee (1963).

On the basis of equilibrium theory one might hypothesize that the greater disparity between student ratings and faculty self-ratings, the greater the likelihood that behavioral changes would occur in instructors. Thus, large differences in student-teacher perceptions would create the greatest amount of imbalance or dissonance in teachers. This is essentially what Centra's monograph (1972b) showed in demonstrating the utility of student ratings for instructional improvement. His major finding was that knowledge of results (feedback) generated improvement for those instructors whose self-estimate of their teaching skill was significantly higher than student ratings.

In addition, empirical data suggested that traditional evaluative practices by supervisors present an impediment to improving instruction. As an example, Tuckman and Oliver (1968) found that when supervisor feedback was the only information provided to teachers, it resulted in a significant but negative change in teacher behavior. In that study teachers altered their behavior in the opposite direction to that suggested by their administrators. On the other hand, when student feedback was the only information provided, it resulted in a significant positive effect on teacher behavior. Apparently, teachers feel

threatened by supervisors who employ vague or irrelevant criteria and, consequently, reject their supervisor's advice (Good and Brophy, 1973). Therefore, student feedback, in contrast to administrator feedback, appears to exert a profound, positive impact on subsequent teaching performance.

Evaluating Teaching

A primary aspect of any teacher evaluation is to determine a value judgment (Ryans, 1967; Ahmann, 1972). In the past, quantity of research publications have been the major criterion (value judgment) for evaluating college teachers. Because of this emphasis on research, Eble (1971) suggested that the majority of university faculty are indifferent about teaching and are not concerned with improving their teaching skills. Compounding this problem, there is little or no correlation between amount of research published and student ratings of teaching effectiveness (Costin et al., 1971). In addition, King (1972) believed that many instructors hide behind tenure without producing in the classroom and that most administrators ignore this fact in order to avoid a controversial lawsuit. Bertramson (1972) championed the idea that administrators must demonstrate behaviorally that meritorious teaching is rewarded equally with research merit.

As an American custom, research productivity has outweighed teaching productivity as the main criterion for professional advancement in the majority of the larger colleges and universities (Hildebrand et al., 1971). It appears that this traditional practice has been conducted over the years in part, at least, because reliable and

particularly, valid instruments for evaluating teaching were scarce or nonexistent. But Jencks and Riesman (1968) offered a different explanation. They mentioned that the history of student evaluations of faculty in American higher education revealed that teachers will research and study almost everything under the sun--except one thing--themselves. Jencks and Riesman (1968) stated that faculty have discouraged student ratings under the protective aegis of the democratic processes such as faculty senates or other governing bodies of colleges. As one recent example, the faculty senate at the University of California at the Davis campus defeated a proposal to use student ratings to evaluate teaching (Hildebrand, 1972). In this respect, Sord (1973, p. 62) believed that such an overemphasis on research competence in contrast to teaching competence reflects a serious educational threat which is a "little short of scandalous." Metcalfe (1972, p. 42) questioned whether research can be so easily evaluated: "I don't see many questionnaires for the evaluation of research floating around. Research is not infallible. . .much of the research is still evaluated by the number of pages in the scientific journal."

The net effect of this overemphasis on research productivity in contrast to teaching productivity results in faculty investing their time and energy in research activities instead of in teaching. It goes without saying that the whole educational environment is diminished by this customary evaluation practice called the "publish or perish" game. Such overemphasis on research output may have arisen, to some degree, from the complexity of the whole teaching-learning, interpersonal

process. It is threatening to many instructors' self-esteem to have their teaching evaluated. For many faculty just the word "evaluation" is a threatening process--what Wilhelms (1967) suggested has been referred to as the "sickman of education."

Any evaluation procedure has certain inherent limitations and perhaps some negative aspects. However, the end result of not having a teacher evaluation program tends to lead to an insidious, "hidden agenda" evaluative procedure (Wilhelms, 1967). By inadvertently committing a school or college to a program of nonevaluation, this irresponsible action tends to diminish or impede certain teaching-learning potentialities. Metcalfe (1972) stated that there is no way to avoid evaluating faculty. When a new teacher is hired, promoted, or assigned to various committees, he is evaluated. The question is not whether teachers should be evaluated but rather how the evaluation can be conducted most fairly and accurately. Wilhelms (1967) believed that the very distinction of being human and imperfect rests on the "need for" and "potential for" being evaluated. Similarly, Metcalfe (1972, p. 35) echoed the following conviction that people in all kinds of gainful work are evaluated: "Why not evaluate those who teach? The professor can no longer play the role of 'God' in the classroom. . . He has felt that he was inviolable and couldn't be touched."

Whether informally or formally, faculty have been evaluated in institutions of higher learning for years. Some questions and controversial issues still remain today. That teacher evaluation is continuous and inescapable on every campus is a postulate that few people would

refute. We can no longer afford to ask, "Should teaching be evaluated?" Instead, the pertinent question becomes, "Do we have reliable, valid, and systematic ways for making the evaluations?"

Research Related to the Psychometric Qualities of Rating Instruments

Many educators do not believe college students are competent to evaluate their teaching on a reliable, valid basis. However, numerous investigations on teaching effectiveness tend to refute this erroneous, misinformed view. In a comprehensive summary of the research on college teaching, Costin et al. (1971) and Centra (1972a) presented empirical findings indicating high reliability and acceptable validity of student ratings.

Reliability

There are empirical answers to many of the questions concerning the reliability of student ratings of instructors. Ratings are made with high reliability, especially if there are 20 or more raters (Hoyt, 1969). Ratings are generally as reliable as the better mental tests available (Remmers, 1960). In one study, alumni ten years after graduation agreed (correlations between .40 to .68) with on-campus students in their ratings of the same teachers (Drucker and Remmers, 1951); stability over both short and long periods of time have been reported in the ratings by alumni and by current students of the same teachers (Gage, 1963). Student ratings generally show negligible correlations with class size, class level, required versus elective course,

course in major or not, sex of respondent, grade point average, and expected grade in course (Hildebrand et al., 1971).

Some notable exceptions to these findings have been reported. For example, Hoyt (1969) found that student ratings varied as a function of class size and specific instructional objective (item on which teachers were rated). Rank and years experience in teaching correlated significantly with student ratings (Aleamoni and Spencer, 1971). Costin et al. (1971) believed that some course characteristics (elective, major, or minor) influence student ratings of instructors. In general, it has been found that student and course characteristics do not significantly bias student ratings of teachers (Hildebrand et al., 1971).

Validity

The following studies focused on the validity issue by reviewing the relationships between student ratings of effective teachers that seemed to be consistent with criteria of effective teaching (Costin et al., 1971; Centra, 1972a). For example, students in an aircraft mechanics course made gains in information and in practical "job sample" performance which correlated significantly with overall ratings of the course (Morsh, Burgess, and Smith, 1956). In another study, Elliot (1950) found a low but significant positive relationship between student ratings of chemistry teachers and their achievement in the subject. Also, students with higher achievement scores in introductory psychology courses rated more favorably the overall course in contrast to students with lower achievement (Russell and Bendig, 1953).

In several other studies, students showed that they learned (had greater achievement gains) from teachers whom they said gave clear explanations, were attentive, friendly, and flexible. In some instances, student achievement was influenced by the interaction of student and faculty characteristics. McKeachie, Linn, and Mann (1971) reported that students high in need affiliation achieved best in critical thinking with instructors they perceived and rated as being warm and sympathetic. In addition, several factor analytical studies have been performed on student ratings of instruction. From these studies certain common factors or clusters emerged with similar overlapping results: (1) organization or preparation, (2) teacher skill, (3) stimulation, (4) teacher-student rapport, and (5) clarity of communication (Hildebrand et al.; 1971; Centra, 1972a).

The Content of Rating Scales Used By Previous Researchers

Most researchers have developed rating items from four basic domains: (1) teacher personality characteristics, (2) teaching skill, (3) organizational structure of the course, and (4) teacher-student interactions.

Teacher Personality Characteristics

Teacher Warmth. Some teacher personality dimensions have been found to correlate substantially with effectiveness in teaching (Gage, 1972). An important facet of personality is the affective domain consisting of attitudes, feelings, and emotions. Ryans (1960) and

McKeachie (1969) generated rating items dealing with the instructors' sensitivity to student needs and feelings. Flanders (1960) and Rogers (1969) presented evidence that student achievement increases when teachers maintain rapport, are warm, personable, and encouraging. Kerlinger (1966) and Hamachek (1968) cited evidence documenting the importance of teacher warmth. Typical rating items showing this teaching quality are as follows: (1) "In this class I felt free to ask questions or express opinions" (Centra, 1972); and (2) "Was sensitive to student feelings and problems" (McKeachie, 1969).

Stimulation. Ability to generate intellectual excitement and intrinsic motivation for learning appears to have its benefits. For example, Ryans (1960) found that teachers who are lively, stimulating, and enthusiastic about their subject are judged to be more successful by administrators and other experienced observers of teaching. In addition, students made greater reading comprehension scores when taught by enthusiastic teachers. In almost every major factor analytic study on teaching, stimulation is one common element consistently reported (Eble, 1971). Some typical rating items from this category are the following: (1) "Is enthusiastic about the subject" (Hildebrand et al., 1971); and (2) "It was easy to remain attentive" (Aleamoni and Spencer, 1971).

Teacher Model (Example). One salient facet of the teaching-learning process involves the impact of models on students. Both verbal and nonverbal communications provide an important source of interaction in which learning is enhanced by the imitation of models. The significant effect of modeling on the behavior of students has been well

documented (Gage, 1963). Bandura and Walters (1963) and Bandura (1969) substantiated that a warm, supportive relationship between the adult model and student enhances learning. McDonald and Allen (1967) presented evidence demonstrating the positive effects of modeling in changing teacher behavior. Good and Brophy (1973) summarized relevant research on the teacher as a model. Langen (1966) constructed items related to the impact of the teacher as an exemplary figure: (1) "Inspires class confidence in his knowledge of the subject," and (2) "Has motivated me to do my best work."

Teaching Skill

Open Communication. Instructors who are highly effective are willing to discuss points of view other than their own. Hunt and Joyce (1967) found that effective teachers tend to encourage dialogue and meaningful discussion. Amidon and Hough (1967) presented comprehensive evidence showing that effective teachers enable students to ask questions and express diverse views and feelings in an open classroom climate. Several representative rating items which help illustrate this teaching quality are these: (1) "Encouraged students to express themselves freely and openly" (Hoyt, 1969); and (2) "The instructor was open to other viewpoints" (Centra, 1972a).

Subject Matter Knowledge. There is at present no adequate student measure of a teacher's grasp of his subject matter field (Hoyt, 1969). However, it seems evident that a teacher cannot furnish adequate feedback to students or clarify ambiguities and misconceptions unless he has an adequate understanding of the subject he teaches. The

learning of facts by students has been found to be significantly related to clarity and expressiveness in the teacher (Solomon et al., 1964). Also, there is evidence that ideational fluency of teachers correlates highly with ratings of teacher effectiveness (Knoell, 1953). Aleamoni and Spencer (1971) mentioned one typical item frequently found on student evaluation forms: "The instructor had a thorough knowledge of his subject matter." Hoyt (1969) used this item in his research: "The instructor demonstrated the importance and significance of his subject matter."

Organizational Structure of the Course

Preparation. Research indicates that orderliness facilitates certain learnings. For example, Ryans (1960) reported that "superior" teachers tend to be systematic, highly responsible, and well-organized. Spaulding (1963) found a positive relationship between teacher orderliness and student reading achievement. Two frequently cited items on rating forms are as follows: (1) "Is well prepared" (Hildebrand et al., 1971) and (2) "Used class time well" (Centra, 1972b).

Flexibility. McKeachie (1969) suggested that a well organized course needs to be balanced with a proper amount of teacher adaptation and change. Flanders (1967b) found that student achievement was accelerated when teachers changed their teaching strategies to meet student needs or new situations. Typical rating items are these: (1) "Was he flexible?" (McKeachie, 1969); and (2) "Changed his approach to meet new situations" (Hoyt, 1969).

Teacher-Student Interactions

Flanders (1967a) based a unified theory of effective teaching and learning on what he called "interaction analysis." Hoyt (1969) found that active student involvement was an important factor that contributed to effective teaching. McKeachie (1970) reported that the quality of the teacher-student relationship adds a significant dimension to the educational process. Some commonly used items are the following:

(1) "Invites students to share their knowledge and experience" (Hildebrand et al., 1971); and (2) "There was not enough student participation for this type of course" (Aleamoni and Spencer, 1971).

Research Findings and Theoretical Positions Regarding Effective Teaching

Defining Effective Teaching

It is evident that no two people perceive effectiveness in teaching from the same goal orientation, perspective, and frame of reference. In considering various ways to describe and characterize effective teaching, definitions of teaching have varied greatly. Teaching is perceived as a subjective art by some people and as an objective science by others. In this respect, Flanders (1967b, p. 235) asserted: "There are those who say that teaching is an art; certainly no one can deny that an exceptionally fine lesson has much in common with an artistic performance. Most artists would be the first to point out, however, that arduous, lengthy practice and attention to technical skill were prerequisite to a particular fine performance." McKeachie (1967)

mentioned that teaching, like art, involves complex value judgments. He recognized that many variables interact in determining teaching effectiveness. Combs (1964) voiced skepticism concerning the possibility of ever finding specific, objective criteria of teaching effectiveness. As opposed to Combs' skeptical feelings in this regard, Flanders (1967b) and Gage (1972) believed that researchers are closer than ever before to a scientific understanding of effective teaching.

Flanders (1967b) maintained that it is easier to identify poor teaching than it is to identify effective teaching. The characteristics of the former are more consistent and predictable than the latter. Eble (1971, p. 10) felt that we know a good deal about effective and ineffective teaching:

The profession does have substantial information about the effectiveness of both teachers and teaching. There is no great disparity between the findings of one research study and another, between what research and informed opinion have to say about effective teaching and (at the point of teaching performance) between what students and faculty say.

On the other hand, there are those educators and persons who believe that any effort to depict effective teaching is distorted by the instructor's personal charisma, "salesmanship," or popularity. Hildebrand et al. (1971) discovered that there is substantial agreement among students, and between faculty and students, about the effectiveness of given teachers. These researchers found that best and worst teachers engage in the same professional activities by spending their time among academic pursuits in about the same ways. Therefore, the mere performance of teaching activities does not in itself assure that instruction is effective.

As soon as one attempts to define or delineate effective teaching, the question of a criterion of effectiveness usually enters the picture also. From a cross-cultural perspective, Fishman (1967) mentioned that from the research that is available there is no one type of teacher, teaching, or classroom organization which produces the "best" results with all students. Sord (1973) indicated that the term effectiveness as used in the phrase "teaching effectiveness" does not represent a single, isolated quality, but embraces a number of different and interdependent aspects of excellence, each of which contributes to the final product of effective teaching. Weaver (1966) stated that before one can decide whether a teacher is good, one must ask, "Good for what?" He mentioned that the purposes and objectives of the course and of the teacher need to be considered. Gage (1972, p. 123) added this perspective: "We must no longer be misled by the notion that, because there is one word called teaching, there is one, single, overall criterion of effectiveness in teaching that will take essentially the same form wherever teaching occurs."

Any attempt to define accurately or to assess effective teaching appears fraught with complex difficulties. As an example, Glaser (1963) and Popham and Baker (1970) advocated the use of pre- to post-gains of students as a valid procedure to assess teaching effectiveness. However, this approach at measuring the quantity of learning has considerable methodological flaws and shortcomings in which intervening variables contaminate the results. In addition, Eble (1971, p. 14) stated: "Measuring teaching effectiveness by the 'before and after' achievements

of students would necessitate relying upon measurements of students' abilities and achievements which are themselves under question."

Gage (1972) presented a similar view by pointing out the difficulty in measuring and controlling the complex, elusive, learning variable by means of comparing results from pre- and post-tests. Thus, for learning outcomes, it is virtually impossible to control adequately this "change in behavior" variable. However, awareness of the great difficulty with controlling the elusive learning variable does not mean that researchers should forsake any scientific attempt to measure effective teaching (Gage, 1972).

A Possible Criterion of Teaching Effectiveness

Student ratings are believed to be the most accurate and fairest procedure available to evaluate teaching (Costin et al., 1971; Centra, 1973). Stake (1972) believed that such ratings can indicate the degree of satisfaction experienced from the teaching provided. And Gage (1972, p. 28) referred to the same idea when he mentioned a "something desirable" criterion: ". . .this might be the achievement of students or a favorable evaluation of the teacher by students." Hildebrand et al. (1971) advocated a criterion which they called the "constructive contribution" made to the life of the learner. Probably one of the most comprehensive, all inclusive, statements on this complex topic has been carefully delineated by this research group at the University of California. Hildebrand et al. (1971) illustrated the methodological problems involved with trying to measure progress made by learners toward

the accomplishment of educational objectives by means of a pre-post test comparison. Hildebrand et al. (1971, p. 35) stated: "Facts learned from teachers can be tested, but their value cannot; the contribution a teacher makes to spiritual or emotional maturation cannot easily be assessed." The authors continued by saying that effective teaching must be weighed in light of the constructive contribution the instructor makes to the learner. Such a contribution could be knowledge imparted, counsel given, objectives clarified, human values or skills developed, and professional experience shared. The California study mentioned that the learner may not be able to assess fully the constructive contribution made to his life by a teacher, and student judgment may change with the passage of time. Nevertheless, the student is usually the best judge of the positive contributions made to his own life (Hildebrand et al., 1971).

A Tentative Definition of Effective Teachers

In summarizing the research cited in this chapter on effective teaching, the following is a partial delineation of effective teachers. The effective teacher engages students in a warm, open, encouraging relationship by his willingness to listen and accept their different views and feelings. He enables students to examine their own thoughts, beliefs, and value system. He is an exemplary model of intellectual stimulation and enthusiasm that is contagious. Typically, he knows his subject matter well by clearly communicating it to students. In addition, he is skilled in the use of different planning, facilitative,

and organizational procedures and is knowledgeable about their limitations. Furthermore, he is schooled in various alternative ways to provide and adapt appropriate learning conditions and experiences. Finally, he attempts to have students become actively involved with applying course material through class discussions and other means of student participation.

From earlier sections of this chapter one can see that informal student and administrator evaluations are neither systematic nor objective. Such unsystematic evaluations can be quite inaccurate and unfair. Student volunteers and department chairmen or deans usually are not representative of the group. Open-ended essay responses and reports consume a great deal of time and pose complex difficulties for comparative interpretations. Informal evaluative procedures usually are based on replies from a few students or administrators who may try more for "wise" rather than valid evaluations. Their motivations and purposes can be questioned (Hess, 1972). In addition, the accuracy and reliability of such informal, unsystematic evaluations can be challenged on psychometric grounds. In short, vague, subjective student and administrator opinions cannot be accepted as a substitute for objective evidence of teaching effectiveness (Aleamoni and Spencer, 1971).

Assessing teaching effectiveness by the use of student ratings involves the selection of an appropriate set of items on which value judgments will be made. Carefully selected and constructed items permit a comparison that might be used for normative purposes. With these methodological issues considered, 227 items from five factor analytical

studies on teaching effectiveness served as an empirical basis in the development of the evaluation form which will be described at length in the next chapter.

Summary

In this chapter a historical and empirical perspective on the evaluation of teaching was presented. Informal student and administrative evaluations plus quantity of research publications have dominated procedures used to evaluate teaching in America. Such evaluations have certain inherent weaknesses in that they are largely unsystematic, unreliable, and invalid. Research was cited, demonstrating that student ratings can be used for the improvement of instruction.

From the research studies presented herein, student ratings have high reliability and acceptable validity for evaluating teaching. Thus, student ratings appear as "qualified" as any defensible measure to assess teaching effectiveness. Also, in this chapter the value of the "constructive contribution" to the life of the learner was mentioned as a possible criterion which might be used to evaluate teaching. Lastly, on the basis of a review of the literature on teaching effectiveness some of the demarcating qualities of effective teachers were summarized.

CHAPTER 3

METHODOLOGY

This chapter will contain a description of samples used in the study. Also, it will describe the development of instruments, specific procedures used in administering instruments, and statistical procedures employed in analyzing responses from these instruments. For the sake of convenience, the chapter will be divided into the following sections: (1) obtained samples, (2) development of instrument, (3) administration of instrument and data collection, and (4) scheme for data analysis.

Samples

Teachers

The sample of teachers to be evaluated consisted of 50 volunteers from a total of 140 full and part time faculty in the College of Education. This volunteer group constituted approximately 35 percent of the total faculty population in the College of Education at The University of Arizona. These volunteer teachers belonged to various departments and their departmental affiliation is represented by Table 1. The sampled teachers taught graduate as well as undergraduate classes. All of the faculty members in the College of Education were invited to participate in the project. These volunteer faculty (N=50) constituted the sample of teachers whose classes were asked to evaluate them using the instrument which was developed for this purpose.

TABLE 1
DEPARTMENTAL AFFILIATION OF SAMPLED TEACHERS

Department	Number Belonging to each Department
1. Counseling and Guidance	5
2. Educational Foundations and Administration	5
3. Educational Psychology	9
4. Elementary Education	9
5. Secondary Education	8
6. Special Education	2
7. Reading	3
8. Rehabilitation	7
9. Library Science	2

Students

Students were sampled as a class unit. These sampled classes were used to evaluate the teachers who volunteered for the study. Class size ranged from 5 to 52 students with a mean of 22.5 students per class. Using the final form of the development instrument, 1,650 students rated the degree of effectiveness of the 50 volunteer faculty in 76 of their classes. Eighteen of these classes had students ranging from 5 to 15, 44 from 16 to 30, 10 from 31 to 45, and 4 from 46 to 60. These data are supplied in Table 2.

TABLE 2
NUMBER AND SIZES OF SAMPLED CLASSES

Class Size	Number of Classes per Category
Small (5-15)	18
Small-Medium (16-30)	44
Medium-Large (31-45)	10
Large (46-60)	4

Development of Instrument

Formation of Committee on the Evaluation
of Teaching

The Committee on the Evaluation of Teaching consisted of six experienced faculty members who were appointed by the Dean of the College of Education to assist the author of this dissertation in the administration of this project. Another function which the committee assumed was to help establish the goals and objectives of the project on evaluating teaching effectiveness and also to make recommendations for selecting appropriate items consistent with the project's objectives. The background and experience of committee members ranged from their abilities in psychometrics, school and university teaching to administrative competencies in education. This committee also served as a group of experts to at least determine the face validity of the items included in the final version of the instrument. They represented their departmental concerns and philosophies which were of tremendous value in coming up with a unique instrument capable of evaluating the

effectiveness of teachers using diversified teaching strategies. Their departmental affiliation and area of expertise are presented in Table 3.

TABLE 3
TEACHER EVALUATION COMMITTEE MEMBERS' DEPARTMENTAL
AFFILIATIONS AND COMPETENCIES

Departmental Affiliation	Number on Committee	Committee Members' Special Competencies
Special Education	1	Behavior modification and learning disorders
Educational Psychology	1	Measurement, research design, and statistics
Elementary Education	1	Elementary science teaching
Secondary Education	1	Educational administration, supervision student teachers
Rehabilitation	1	Counseling and research
College Administration	1	Educational Administration, social psychology and attitude scaling

Selection of Items for the First Draft of the Instrument

In consultation with the Committee on the Evaluation of Teaching in the College of Education during the fall semester of 1972, a 69-item instrument (Appendix A) was constructed. A questionnaire, rating approach suggested by Remmers (1963) served as a guiding rationale. An attempt was made to develop an evaluation form consistent with the unique imprints and philosophical orientation of the College of Education at The University of Arizona. Five instruments developed by other

researchers (Hoyt, 1969; Aleamoni and Spencer, 1971; Hildebrand et al., 1971; Centra, 1972b; and Hartley and Hogan, 1972) served as an item domain for the present study. Taken as a unit, this item domain consisted of a total of 227 items which were submitted to the members of the Teaching Evaluation Committee. The committee members were requested to consider this item pool for selecting items which after modification, if needed, could be used for the initial try-out phase. The committee, thus, selected only items for this pool which indicated a high face validity and were believed to be related to instructional effectiveness.

In addition to selecting items for this pool, an effort was made to write additional items based on the following guiding criteria: (1) each item presented a meaningful instructional dimension; (2) each item showed research evidence of being highly related to instructional effectiveness; (3) each item described teacher behavior in sufficiently specific terms that, if it seemed desirable to modify the teacher activity, this could be conveyed clearly via knowledge of results; (4) each item's concept appeared at least two or more times in the total pool of 227 items from the five factor analytical studies; (5) each item tended to be associated with the higher cognitive and/or affective levels of Bloom's Taxonomy of Educational Objectives; and (6) each item was known to be associated with sound pedagogical practices.

Utilizing the above criteria, the committee members by consensus recommended 49 items which constituted the first draft of the instrument. In addition, on the recommendation of the consulting committee a set of ten additional items were selected and constructed to describe

student reaction to general aspects of the course and instructor (such as assigned readings, quality of exams, overall course value, teacher recommendation, etc.). Also, based on various investigations of effective teaching and learning, the following three items (12, 22, and 31) were constructed and unanimously approved by the committee: (1) "Provided feedback within a reasonable time period on examination results or other academic work;" (2) "Made a significant impact on my outlook, philosophy, or style of life;" and (3) "Provided students with an opportunity to participate actively in some decision-making processes directly related to the course."

For depicting course-student characteristics, seven additional items were constructed dealing with certain variables like class size, class level, expected grade in course, sex, grade point average, reason for course selection, and required or elective course. These seven additional items were intended to provide valuable feedback information to individual instructors and/or to be used for possible research purposes. Again, the committee unanimously sanctioned the entry of these items.

By carrying out the above procedures, the 69-item instrument (Appendix A) was constructed. In brief, items 1-53 characterized effective teaching procedures and items 54-69 depicted student characteristics and overall course ratings.

Description of the Initial Instrument

The items used in this instrument were confined to recurrent dimensions of effective teaching that have been frequently reported by numerous researchers. As an example, at least seven factors were

reported as recurrent "clusters" or themes in the five factor analytical studies (Hoyt, 1969; Aleamoni and Spencer, 1971; Hildebrand et al., 1971; Centra, 1972b; and Hartley and Hogan, 1972) which served as an empirical foundation for this investigation. These seven common clusters or factors played a key part as a guiding impetus in the selection of items to be constructed for the 69-item instrument. The seven hypothetical factors are presented in Table 4 with their respective logical item elements that comprised the major body of the initial instrument.

Table 4 depicts the essential items (1-53) that are related to effective teaching activities of the 69-item instrument. Of further importance are those items (54-69 in Appendix A) which were constructed to obtain information about student characteristics and ratings related to certain course activities. It was felt that these additional items such as instructor recommendations (54 and 55), expected grade (56), class level (57), sex (58), grade point average (59), class size (60), reason for course selection (61), and course description (62) would provide relevant information for understanding the particular assets and limitations of the developed instrument. Finally, items 63-69 described students' overall ratings of particular aspects of the course such as assigned readings (63), quality of exams (64), quality of lectures or class presentations (65), value of class discussions (66), labs, workshops, or seminars (67), textbook (68), and course value (69).

TABLE 4
ITEMS COMPRISING SEVEN HYPOTHETICAL FACTORS
OF THE 69-ITEM INSTRUMENT

- I. PREPARATION AND ORGANIZATION
 - 4. Objectives clearly stated
 - 8. Class time used well
 - 10. Accomplished stated objectives
 - 14. Well informed
 - 18. Well prepared
 - 21. Covered appropriate number of topics

- II. STUDENT INVOLVEMENT (PERSONALISM)
 - 9. Readily available for individual help
 - 13. Encouraged student speaking
 - 17. Facilitated class discussion
 - 24. Genuinely interested, actively helpful
 - 31. Opportunity for active participation
 - 37. Sensitive to student feelings

- III. CLARITY OF COMMUNICATION
 - 2. Answered student questions completely
 - 3. Sufficient examples to clarity
 - 5. Summarized or stressed major points
 - 11. Explained clearly
 - 20. Showed concern with student confusion
 - 29. Distinguished significant-insignificant material
 - 35. Spoke with expressiveness-variety

- IV. STIMULATION
 - 1. Enthusiastic about teaching subject
 - 6. Put forth best effort in learning
 - 16. Independent thinking, reasoning, etc.
 - 23. Stimulated interest in his talk
 - 30. Stimulated thinking by raising issues
 - 34. Increased student interest in course material

- V. TEACHING STYLE (SKILLS AND ACTIVITIES)
 - 7. Discussed other viewpoints
 - 19. Was flexible, changing approach
 - 25. Students applied concepts
 - 27. Presented material in humorous ways
 - 32. Made good use of teaching aids
 - 38. Showed self confidence in subject
 - 39. Model of professional growth

TABLE 4--Continued

-
-
- 40. Encouraged and praised students
 - 41. Enabled students to answer own questions

VI. EXAMS AND EVALUATION

- 12. Gave feedback within reasonable time
- 15. Exams reflected objectives
- 26. Gave helpful feedback on papers
- 28. Told students how they are evaluated
- 33. Assigned grades on fair reasonable basis
- 36. Gave conceptual exams
- 42. Gave fair and appropriate assignments

VII. STUDENT LEARNING (SELF-EVALUATION)

- 48. Principles, generalizations, etc.
 - 49. Broad liberal education
 - 50. Implications for self growth
 - 51. Oral or written self-expression
 - 52. Sense of personal responsibility
 - 53. Creative problem solving
-

Administering Initial Instrument

In the fall semester the 69-item instrument was tried out to evaluate 50 faculty volunteers in 82 or their classes which consisted of approximately 1,900 students. These teachers were asked to administer the instrument to their own class(es) and asked to step out of the room during the ratings in an attempt to maximize their reliability. Additionally, the anonymity of instructors was maintained by their self-selection of a random five-digit number for maintaining confidentiality in communicating the results. By the same token, student anonymity was safeguarded by requesting that their names not be given on the IBM answer sheets. For students using the 69-item form, specific directions were given them to read before rating their instructors. These directions appeared on the front page of the instrument (Appendix A).

Selection of Items for the Final Draft of the Instrument

One of the major problems with student ratings of instruction results from discrepant goals and expectations of students and faculty (McKeachie et al., 1971). In consideration of this problem an effort was made to obtain validity indices from both students and faculty members. Therefore, a 52-item shortened version (Appendix B) of the 69-item instrument was used for the purpose of assessing the degree of agreement between faculty and student perceptions of what constitutes effective teaching. These 52 items (leaving 17 items intended to obtain student-course related information), thus selected, were given again to the same group of students to determine the appropriateness of these

items for evaluating the effectiveness of teachers. This rating scale also was administered to a group of participating faculty. A six point, Likert-type, rating scale was used to rate each item by students as well as by faculty. One, on this scale, represented the unideal item for evaluating teaching effectiveness and six, the point on the other extreme, indicated an ideal item for evaluating teachers.

This 52-item intermediate instrument, with specific instructions, was submitted to faculty and students to determine those items which would be commonly rated by the two groups as the most appropriate ones that might be included in the final version of the 47-item instrument. Thus, the final version of the developed instrument contained those items from the 52 items which were mutually rated as appropriate ones for evaluating teaching effectiveness. Similarly, items rated as unideal or inappropriate were discarded.

Administration of Instrument and Data Collection

Administration of the 52-Item Instrument

In the fall semester, all 140 full and part time faculty in the College of Education were mailed this 52-item instrument. Eighty-six instructors returned the completed instrument. This represented a 61 percent return rate.

Likewise, for this phase, using the 52-item instrument, 169 students indicated their relative preferences for certain ideal, effective teaching procedures. All 169 students were considered to be representative of the student body of the College of Education for the

fall semester of 1972. This student group was obtained from various levels of required classes such as educational psychology courses for undergraduate juniors and seniors, and research courses for graduate students at the master's and doctoral levels. This student sample was composed of 83 undergraduates, 71 master's, and 15 doctoral level students who filled out the instrument in their respective classrooms. This student sample in this instance, therefore, was representative of those faculty who taught either undergraduate educational psychology or graduate research courses and allowed their students to participate in this intermediate phase of the study.

Description of the Final Instrument

Using the criteria of a high degree of agreement between faculty and student ratings, 32 items from the 52-item instrument were selected for the construction of the final rating scale. Of these 32 items, thus selected, 27 items represented again the item domain of the seven hypothetical categories (Table 4) of the initial 69-item instrument. These 27 items are presented in Table 5 to show the continuity of the instrument's development.

In the construction of a final instrument (Appendix C), 15 additional items were added to the 32 items thus selected, making this instrument 47 items long. The first eight items (1-8) on this developed instrument furnished feedback concerning student-course characteristics. The next 32 items (9-40) were to determine instructional effectiveness, and the last seven items (41-47) provided information about the students' global reactions toward the instructor and the course.

TABLE 5

ITEMS COMPRISING SEVEN HYPOTHETICAL FACTORS OF THE 47-ITEM INSTRUMENT

-
-
- I. PREPARATION AND ORGANIZATION
 - 19. Class time used well
 - 21. Well prepared
 - 23. Accomplished objectives
 - 35. Objectives clearly conveyed

 - II. STUDENT INVOLVEMENT (PERSONALISM)
 - 24. Facilitated class discussion
 - 30. Opportunity for active participation

 - III. CLARITY OF COMMUNICATION
 - 15. Clearly explained course material
 - 17. Summarized or stressed major points

 - IV. STIMULATION
 - 16. Student effort in learning maximized
 - 27. Stimulated student interest and attention
 - 34. Stimulated thinking by raising issues

 - V. TEACHING STYLE (SKILLS AND ACTIVITIES)
 - 14. Willing to discuss other viewpoints
 - 20. Student opportunity to apply concepts
 - 22. Was flexible, willing to change approach
 - 26. Enabled students to answer own questions
 - 29. Encouraged and praised students

 - VI. EXAMS AND EVALUATION
 - 18. Gave helpful feedback on exams or papers
 - 25. Provided feedback within a reasonable time
 - 28. Exams or other work consistent with objectives
 - 31. Gave exams or work which were conceptual
 - 32. Assigned grades on fair, reasonable basis
 - 33. Told students how they are evaluated.

 - VII. LEARNING (SELF-EVALUATION)
 - 9. Developed or learned specific skills
 - 10. Developed or exercised creative problem solving
 - 11. Discovered or learned implications for self growth
 - 12. Developed a sense of personal responsibility
 - 13. Exercised skills of oral or written self-expression
-

Administration of the Developed Instrument

During the 1973 spring semester all full and part time faculty in the College of Education were mailed a letter (Appendix D) by the project director of the Committee on the Evaluation of Teaching. Volunteers were requested to participate in this evaluation project for the second phase of the instrument's development. Faculty members who volunteered for the study were assured that the results of their individual rating reports would remain confidential and that their anonymity would be safeguarded. This was done by having faculty select any random five-digit code number.

Furthermore, a standard procedure (Appendix E) for administering the forms was recommended to the faculty volunteers. All participating faculty members were given specific directions for distributing the rating scale to their classes and for returning the completed ratings later on to the author. It was suggested that instructors absent themselves from the class while students were rating them. This policy was suggested to insure student anonymity and to maintain the reliability of the ratings. In addition, specific directions for students were printed on the front page of the instrument (Appendix C).

A Scheme of Evaluation Based on Teachers' Preferences

In the past, rating scales, used for evaluating instruction, have been criticized for not having teachers determine their own instructional goals and generating their own items for being evaluated (Centra, 1973; Frey, 1973). Consistent with such a suggestion, the present study asked

participating faculty members to select items which seemed completely congruent with their teaching strategy(ies). In order to accomplish this objective each instructor was requested to identify the items on which they preferred to be evaluated (Appendix F). Such a preference was indicated by each faculty member by circling the items on the final form. Students, however, were not made aware of these faculty preferences and used all the items in rating their instructors. This methodology, thus, facilitated the analysis of the responses on preferred and non-preferred items of the instrument. It should also be noted that for indicating faculty preferences only those items (9-35) were used which were designed on an equal interval, five-point, Likert-type scale to evaluate teaching.

Scheme for Data Analysis

Analysis for Determining Psychometric Qualities of the Instrument

Three statistical analysis techniques were utilized for analyzing the responses of the subjects, for determining the item characteristics, and for testing specific hypotheses. These techniques are as follows: (1) use of mean and standard deviation and t tests, (2) analysis of variance, and (3) correlational procedures.

Means and standard deviations were used as the criteria for selecting the items for the final form. By using the 52-item intermediate instrument, in general, those items rated highly (5.00 or greater on a 6.00 scale) by both faculty and students were included in the final

statistical analysis performed between these two groups. More exactly, a t test was conducted on each of the 52 items by comparing the means and standard deviations of the faculty and student groups. Those items showing a significant difference at either the .05 or .01 level were discarded from the 69-item form in the revision of the 47-item revised instrument. However, for both practical and theoretical reasons a few crucial items which showed significant differences at these criterion levels were retained and included in the 47-item developed instrument. By employing the above statistical operations in using the 69-item and 52-item instruments, an attempt was made to try to maximize the final 47-item instrument's discrimination power and criterion validity.

In addition to the above procedures, an analysis of variance technique was employed to analyze the responses intended to test two specific hypotheses related to class size and teacher preferences. A two by three ANOVA design was used to accomplish this purpose.

Finally, a correlation coefficient was computed to determine the reliability of the final instrument. For this purpose Cronbach's Alpha (1951) was used.

Hypotheses to Be Tested

Two major hypotheses were formulated for this study. One hypothesis related to the effect of class size on student ratings of teaching effectiveness. Another hypothesis related to the effect of faculty members' unique style of teaching, goals, and objectives on student ratings of instruction. These hypotheses are stated below.

Hypothesis 1. There will be no significant differences in the mean ratings made by the small, medium, and large classes of students in evaluating their instructors.

Hypothesis 2. There will be no significant differences in the mean scores of the faculty "selected" versus "nonselected" items as indicated by student ratings of the volunteer teachers.

In addition to these hypotheses the major purpose of the study was to come up with a desirable, reliable, and valid instrument which could be used by the college for evaluating teaching effectiveness.

Summary

This chapter described the sample of teachers and students used in the development of the 47-item instrument plus specific developmental, administrative, and statistical procedures conducted for this investigation. Two sequential phases of the instrument's development were delineated with their respective item composition presented in tabular form. The chapter also described several methods employed in validating the developed instrument. Finally, two central hypotheses of the study were mentioned.

CHAPTER 4

RESULTS

This chapter will describe results related to the psychometric qualities of the instrument and hypotheses intended to be tested. The psychometric qualities of the final instrument will include: (1) results related to item characteristics, (2) validity data, (3) reliability data, and (4) college norms with the final form. With regard to the hypotheses of the study, the results will be presented for describing the effect of class size and faculty self-selection of items on teaching effectiveness.

Results Related to Item Characteristics

Based on the responses of 1,900 students, means and standard deviations were computed for the first 53 items of the initial, 69-item instrument (Appendix A). These results are summarized in Table 6. As evident in Table 6, items 54-69 were disregarded in the analysis in that these items were designed to obtain data related to student-course characteristics. In addition, these items (54-69) were not constructed in a Likert-type format.

The means and standard deviations were used as criteria for selecting items. The mean ratings of the items ranged from 2.04 to 4.67. It is of interest to note that the mean ratings of most of the items centered around 3.50 or higher and hence were skewed considerably.

TABLE 6
 MEANS AND STANDARD DEVIATIONS OF 52 ITEMS
 OF THE 69-ITEM INITIAL INSTRUMENT

Item No.	Mean	S.D.
1	4.51	0.87
2	4.52	0.80
3	4.30	0.89
4	4.20	1.06
5	4.22	0.98
6	3.79	1.19
7	4.14	0.99
8	3.22	1.12
9	4.32	1.01
10	4.20	1.04
11	4.13	1.04
12	4.17	1.28
13	4.60	0.82
14	4.67	0.74
15	3.67	1.61
16	4.15	1.08
17	3.31	1.14
18	4.23	1.07
19	4.20	1.09
20	4.38	0.94
21	4.30	0.97
22	3.44	1.40
23	3.12	1.17
24	4.37	0.93
25	4.12	1.13
26	3.77	1.42
27	3.65	1.15
28	4.06	1.27
29	3.04	1.23
30	3.90	1.14
31	3.89	1.27
32	3.69	1.36
33	4.08	1.43
34	3.97	1.18
35	4.16	1.15
36	3.20	1.39
37	4.31	1.02
38	4.41	0.92
39	4.06	1.16
40	4.07	1.14
41	4.05	1.04

TABLE 6--Continued

Item No.	Mean	S.D.
42	4.28	1.05
*43	1.57	1.24
44	3.08	0.83
45	3.07	0.90
46	2.04	1.12
47	2.95	0.76
48	4.19	1.06
49	2.87	1.52
50	3.85	1.19
51	3.78	1.29
52	3.77	1.23
53	3.72	1.28

*Item 43 was not scaled on a Likert-type five-point interval and is a repeated expansion of item 7.

Using means and standard deviations as a criteria, if an item mean fell within the ranges of 2.00 - 4.20 and had a standard deviation of approximately 1.00, such an item was considered suitable for inclusion in the final version of the developed instrument. On this basis, then, items 4, 5, 6, 7, 8, 10, 11, 12, 15, 16, 17, 18, 19, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 39, 40, 41, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, and 53 were selected for later use. Items 1, 2, 3, 9, 13, 14, 20, 21, 24, 37, 38, and 42 were excluded for later use in that they did not attain the specified criteria. This procedure of item selection was utilized to maximize item discrimination (Guilford, 1954).

Validity Data

Consistent with the purposes of the study each item designed to measure teaching effectiveness was validated using two criteria. These two criteria consisted of obtaining teacher and student opinions with regard to the appropriateness of each item of the initial instrument. Using such an approach to establish the validity of items, those items indicating high congruence between student and faculty opinions were judged to be valid. Such a procedure for establishing item validity has been highly recommended by Horst (1968).

From the 69-item instrument the first 52 items (Appendix B) designed for determining teaching effectiveness were administered to a specifically selected group of teachers and students for determining the validity of these items. Both of these groups rated these items using a six-point, Likert-type scale. A low point on the scale indicated inappropriateness of an item whereas a high point indicated appropriateness.

A statistical procedure for accomplishing criterion validity was conducted by comparing the item mean ratings of both student and faculty groups by using a t technique. Items rated highly (5.00 or greater on a 6.00 Likert-type scale) by both faculty and students were analyzed using t tests. Items 1, 3, 4, 7, 8, 9, 11, 12, 13, 14, 18, 20, 21, 22, 23, 26, 27, 28, 29, 31, 32, 33, 34, 35, 36, 37, 41, 43, 45, 46, and 47 which met the criteria for t-test analysis are presented in Table 7.

In general, items rated below the 5.00 criterion level by faculty and/or students were not considered appropriate for later use. These items (Nos. 2, 5, 6, 10, 15, 16, 17, 19, 24, 25, 30, 38, 39, 40, 42, 44, 48, 49, 50, 51, and 52) are presented in Table 8.

In general, on the 52-item instrument, items (with means of 5.00 - 6.00) which showed a significant difference at either the .05 or .01 level were discarded from the 69-item form in the revision of the 47-item instrument. With this dual criteria in mind, 28 items were discarded from the 52-item instrument and consequently were deleted from the initial 69-item instrument in the process of developing the 47-item form (Appendix C). However, some significantly different items (12 and 13) were retained to be included in the final form. In addition, items (10, 24, 25, 30, 44, 48, and 51) which did not meet the 5.00 criterion level were also retained in the revised instrument. These nine items were included in the developed instrument because of their theoretical and/or practical importance as potential valuable feedback information for improving instruction. The information obtained from these nine items seemed to be consistent with other research (Popham, 1971; Hartley

TABLE 7

T-RATIOS OBTAINED FROM COMPARING FACULTY AND STUDENT
PERCEPTIONS OF APPROPRIATENESS OF THE ITEMS
ON THE 52-ITEM INTERMEDIATE INSTRUMENT

Item No.	Faculty (N=86)		Students (N=169)		t-ratios
	Mean	S.D.	Mean	S.D.	
1	5.3	1.0	5.5	0.9	1.5
3	5.0	1.1	5.3	1.0	2.1*
4	5.6	0.7	5.4	0.9	1.8
7	5.3	0.9	5.3	1.0	0.0
8	5.5	0.7	5.3	1.1	1.2
9	5.4	0.9	5.2	1.2	1.5
11	5.5	0.7	5.4	1.0	0.9
12	5.6	0.8	5.2	1.2	3.1**
13	5.3	0.8	5.0	1.2	2.3*
14	5.3	1.1	5.2	1.1	0.7
18	5.2	1.0	5.0	1.1	0.7
20	5.1	0.9	5.2	1.2	0.7
21	5.5	0.9	5.3	1.0	1.6
22	5.2	0.9	5.1	1.2	0.7
23	5.3	0.7	5.2	1.1	0.9
26	5.5	0.7	5.3	1.1	1.8
27	5.2	1.0	5.0	1.2	1.5
28	5.2	1.0	5.0	1.1	1.5
29	5.5	0.8	5.5	1.0	0.0
31	5.5	0.7	5.3	1.2	1.6
32	5.3	0.9	5.3	1.1	0.0
33	5.7	0.7	5.3	0.9	3.9**
34	5.7	0.9	5.3	1.1	3.1**
35	5.3	0.8	5.3	1.1	0.0
36	5.0	1.2	5.0	1.4	0.0
37	5.5	0.8	5.5	0.9	0.0
41	5.2	0.9	5.0	1.2	1.5
43	5.2	1.0	5.3	1.1	0.8
45	5.2	1.4	5.1	1.0	0.6
46	5.4	0.9	5.2	1.1	1.6
47	5.1	1.1	5.0	1.1	0.7

* = $p < .05$

** = $p < .01$

TABLE 8

ITEMS NOT MEETING THE 5.00 CRITERION LEVEL
FOR BOTH FACULTY AND STUDENT GROUPS
ON THE 52-ITEM INSTRUMENT

Item No.	Mean	S.D.	Mean	S.D.
2	4.7	1.0	5.1	0.9
5	4.1	1.2	4.7	1.4
6	4.8	0.9	5.0	1.1
10	4.1	1.4	4.6	1.4
15	3.9	1.8	4.4	1.5
16	4.6	1.2	4.4	1.4
17	4.9	1.3	4.8	1.2
19	5.1	1.2	4.9	1.2
24	4.9	1.4	4.8	1.3
25	5.1	1.2	4.9	1.3
30	4.8	1.2	4.5	1.3
38	4.6	1.3	4.6	1.3
39	4.5	1.5	4.3	1.5
40	5.0	1.2	4.9	1.2
42	4.5	1.1	4.7	1.3
44	4.9	1.2	5.0	1.1
48	4.7	1.4	4.5	1.4
49	4.8	1.3	5.0	1.2
50	4.8	1.3	4.8	1.3
51	4.2	1.3	4.2	1.6
52	4.5	1.5	4.3	1.5

and Hogan, 1972) and theoretical positions (Gagné, 1965; Bruner, 1966; Hilgard and Bower, 1966) on effective teaching and learning. A final form of the instrument thus developed is included in Appendix C.

Relationship between Perceived Learning and Student Ratings of their Instructors

McKeachie in his research with student ratings indicated, "Student ratings do have some validity. Teachers rated as effective by students tend to be those teachers whose students learn most" (1969, p. 440). No direct test of this important assumption was conducted in the present investigation. However, an indirect test was attempted. Item 40 (part 1 and 2) of the 47-item instrument was used to measure student learning. Information about learning thus obtained by this item served as a criterion for the validity of the instrument. In order to help validate the developed instrument in this way it was assumed that students would rate favorably their instructors on items 9-35 (dealing with effective teaching) in relationship to their responses on part 1 of item 40: "I would recommend the instructor to other students because of the learning experienced." Table 9 shows the degree to which students rated their instructors as being effective on items 9-35 in relation to their ratings on item 40, part 1 or 2.

The results of a t test conducted to test the difference between the two obtained means for item 40: (1) would recommend instructor because of learning experienced, or (2) would NOT recommend instructor because of learning very little; and their relationship to the responses on items 9-35 are presented in Table 9. Differences were found to be

significant. This shows that teachers who were recommended because of the learning experienced by students received significantly higher ratings on items 9-35 than teachers who were not recommended because students perceived that they learned very little from these instructors. From such results it would appear that an item of this kind would be an appropriate one to include in a student rating instrument of effective teaching.

TABLE 9

T TEST OF FREQUENCY OF RESPONSES FOR ITEM 40 (1 OR 2)
IN RELATIONSHIP TO ITEMS 9-35 COMBINED

Student Ratings of Instructors	N	Mean	S.D.	t ratio
Item 40 (1 recommended)	1450	4.13	.62	26.77**
Item 40 (2 not recom- mended)	189	2.85	.61	

** $p < .01$

Reliability Data

Reliability of the 47-Item Instrument

The reliability of items 9-35, which were designed on a Likert-type five-point scale to measure teaching effectiveness, was determined by using Cronbach's Alpha (Cronbach, 1951). Using this technique, the reliability coefficient was found to be .85.

Intercorrelations of the 47-Item Form

Based on Bloom's Taxonomy of Educational Objectives (Bloom, Krathwohl, and Masia, 1964) concerning both cognitive and affective domains,

items 9-13 were intercorrelated with the overall means of these items. It was predicted that items 9-13 would have high intercorrelations in that they are believed to be related to global aspects of effective teaching (Hoyt, 1969). The data supplied in Table 10 substantiated this assumption. All of the intercorrelations for items 9-13 were found to be significantly above zero ($\rho < .01$).

TABLE 10
INTERCORRELATIONS OF ITEMS 9-13 OF THE 47-ITEM INSTRUMENT

Item No.	9	10	11	12	13
9	1.00	.57	.53	.53	.37
10		1.00	.58	.61	.53
11			1.00	.56	.39
12				1.00	.49
13					1.00

In addition to Table 10 above, Table 11 depicts the results obtained by correlating items 9-13 with the overall means of items 9-35 (dealing with specific teaching activities and qualities). As is evident from these results, items 9-13 correlated significantly ($\rho < .01$) with the overall ratings of an instructor. Such significant correlation coefficients indicated the usefulness of these items in evaluating teaching effectiveness.

It was assumed that ratings of items 9-13 which conveyed students' self-evaluations of the quality of learning would be related significantly to their perceptions (ratings) of their instructors with regard to

TABLE 11
ITEMS 9-13 MEANS CORRELATED WITH MEANS
OF ITEMS 9-35 COMBINED

Item No.	Items 9-35 Combined
9	.49
10	.58
11	.48
12	.50
13	.51

quality of teaching. The data of Table 11 supported adequately this assumption. Thus it could be inferred that students' appraisal of their own learning in class may serve as a criterion for evaluating instruction.

Additionally, responses on items 41-47 were correlated with the total ratings (items 9-35). These items (41-47) were intended to obtain information related to certain classroom variables such as value of assigned readings, quality of examinations, quality of class presentations, class discussions, laboratories and seminars, textbooks, and overall course value. Students were instructed to rate the effectiveness of these variables given on a five-point scale. The correlated data thus obtained are summarized in Table 12.

All of these correlations reported in Table 12 were found to be statistically significant ($\rho < .01$). This data also were found to be consistent with the results of other studies (Aleamoni and Spencer, 1971; Hoyt, 1969) dealing with the relationship between such course-related

TABLE 12
 INTERCORRELATIONS OF ITEM MEANS 41-47 WITH RATINGS
 OF ITEMS 9-35 COMBINED

Item No.	Overall Course Related Ratings	r
41	"Assigned readings" with total score	.35
42	"Quality of exams" with total score	.57
43	"Quality of class presentations" with total score	.66
44	"Value of class discussions" with total score	.61
45	"Laboratories, workshops, or seminars" with total score	.50
46	"Textbook" with total score	.28
47	"Overall value of course" with total score	.71

variables and teaching effectiveness. An emerging trend from this data would lead to the conclusion that quality of instruction will be significantly related to students' perceptions of certain process variables (textbook, exams, etc.).

College Norms with Final Form of Instrument

Class means and standard deviations for each major item relating to effective teaching were computed. Overall college means for each major item (9-35) associated with effective teaching activities were furnished to the faculty volunteers for feedback purposes and for possible normative comparisons. This information is given in Table 13. It is of interest to note that most of these scores are skewed 3.5 or higher. Because of the restricted, small sample (N=50) of volunteer teachers, college norms for each item were believed to be the most appropriate method for providing interpretative information to these teachers.

Results Related to Hypotheses

The first major hypothesis of this study concerned the relationship of class size to student ratings. It was hypothesized that students belonging to large (N=31 or more), medium (N=16-30), and small (N=5-15) classes will not show differences in the ratings of their instructors.

Another hypothesis to be tested related to the problem of instructors selecting the items which reflected their teaching philosophies and instructional objectives. It was hypothesized that if teachers' ratings were compared on selected and nonselected items, no

TABLE 13
ITEM MEANS AND STANDARD DEVIATIONS
OF THE 47-ITEM FORM

Item No.	Mean	S. D.
9	4.09	1.05
10	3.64	1.22
11	3.92	1.12
12	3.66	1.17
13	3.76	1.29
14	4.26	1.08
15	4.02	1.09
16	3.56	1.22
17	4.01	1.10
18	3.50	1.44
19	4.27	1.06
20	3.87	1.26
21	4.28	0.99
22	4.06	1.16
23	4.16	1.05
24	4.30	1.10
25	3.89	1.40
26	3.83	1.16
27	4.07	1.18
28	3.87	1.37
29	3.82	1.27
30	3.78	1.37
31	4.26	1.29
32	4.00	1.43
33	4.07	1.33
34	3.96	1.13
35	4.04	1.19

difference will be noticed. In order to accomplish this goal, teachers were asked to indicate the items which they felt would convey a fair assessment of their instructional talent. However, students were not made aware of this research strategy and rated all of the items.

The data obtained to test these two hypotheses were analyzed by using a 3x2 (class sizes x option of selecting items) analysis of variance. The results of ANOVA, thus performed, are summarized in Table 14.

TABLE 14
ANOVA SUMMARY RELATED TO CLASS SIZE AND
FACULTY SELECTION OF ITEMS

Source	df	M.S.	F-ratio
Class sizes (A)	2	0.7921	2.43
Error (a)	39	0.3269	
Faculty item selection (B)	1	4.6375	24.18*
A X B	2	0.4338	2.21
Error (b)	39	0.1918	

* $p < .01$

As evident from Table 14, the F-ratio with regard to class sizes failed to attain significance at .05 or lower level of significance. Non-significant F-ratio ($F=2.43$, $df=2/39$) lead to the acceptance of the null hypothesis 1. Therefore, instructors of large, medium, and small classes did not differ significantly with regard to ratings of their teaching effectiveness. As seen in Table 15, the mean ratings of instructors belonging to classes of different sizes were found to be similar.

TABLE 15
 MEAN RATINGS ON FACULTY SELECTED AND NONSELECTED ITEMS
 IN RELATIONSHIP TO CLASS SIZE

Class Size	Mean Ratings on Faculty Selected Items	Mean Ratings on Faculty Nonselec- ted Items
Small (N=5-15)	4.00	3.39
Medium (N=16-30)	4.07	3.46
Large (N=31-up)	3.56	3.52
For all class sizes	3.87	3.40

Contrary to hypothesis number 1 (related to class size), the second hypothesis dealing with faculty selection of items was rejected. It was found that teachers were rated significantly higher on the items they selected than on items nonselected ($F=24.18$, $df=1/39$). The mean ratings obtained on preferred items was significantly higher than the mean ratings on non-preferred items.

Additionally, class size and faculty selection of items showed no interaction effect. It is interesting to note that 42 faculty members selected on the average 13 items from the total of 27 items intended to measure teaching effectiveness. This indicated that at least half the items of the instrument were found to be consistent with instructors' general pedagogical convictions.

Summary

This chapter presented empirical data related to the development, validity, and reliability of the final version of the 47-item instrument. Methods for discarding and retaining items in the revision of the developed instrument were described. Item means and standard deviations of the 69-item initial instrument and the 52-item intermediate instrument were used as criteria for the revision of the final 47-item form. Using Cronbach's Alpha (1951), the reliability of the developed instrument was found to .85. Other intercorrelations were computed, showing that the instrument has high internal consistency of its items. In addition, college norms were presented for items 9-35 of the 47-item instrument.

Chapter 4 also presented evidence showing no significant impact of the effect of class size (small, medium, and large) on student ratings of teaching effectiveness. Additionally, data were reported demonstrating that students rated teachers significantly higher on faculty self-selected items in contrast to unselected ones. This confirmation of the higher student ratings in relation to the faculty selected items was considered an idiographic method used to help validate the developed rating instrument.

CHAPTER 5

SUMMARY, CONCLUSIONS, IMPLICATIONS, LIMITATIONS, AND RECOMMENDATIONS

Current literature suggested that procedures for evaluating teaching typically are unsystematic and unobjective. A growing body of research showed that student evaluations can help facilitate the improvement of teaching on a reliable and valid basis. Several studies reported that class size seemed to influence student ratings of teachers. Crucial to any teacher assessment program is a procedure which considers each faculty member's instructional objectives and style of teaching.

Summary

The purpose of this study was to construct an instrument that might be used to measure teaching effectiveness validly and reliably. Six faculty representatives helped determine the content validity of the instrument from a pool of 227 items. On this basis a 69-item instrument was constructed, tested, and revised in a pilot project. In this process the first 52 items depicting effective teaching were submitted to a representative group of faculty and students in an effort to determine each item's criterion validity and potential usefulness. On this 52-item instrument items rated mutually high by both faculty and students were considered for inclusion in the final instrument. On this basis the final 47-item form was developed.

In addition to the development of an instrument to measure teaching effectiveness, this study sought to answer the following questions. Does class size influence student ratings of educators? Are faculty rated significantly higher on items related to their instructional intentions?

The 47-item instrument was finally developed to measure teaching effectiveness. Items 9-35 were constructed for this purpose whereas items (1-8 and 36-47) were used to provide relevant information concerning the students and course. The final form of the instrument thus developed was administered by a sample of 50 volunteer faculty to 1650 students in the College of Education at The University of Arizona. This sample of 50 volunteer teachers was affiliated with nine instructional departments. The statistical analyses performed consisted of correlational analysis, analysis of variance, and t technique. A 3x2 ANOVA design was used to test the two major hypotheses of the study.

Conclusions

The first hypothesis concerning small, medium, and large classes revealed that class size had no effect on student ratings of their instructors. These results are contrary to some findings (Downie, 1952; Gage, 1961; Hoyt, 1969) in which class size was found to be related to instructional efficiency when measured by student ratings.

The second hypothesis concerning faculty selected items demonstrated significantly higher mean ratings on items which were compatible with teachers' objectives. The data showed that teachers in general were rated significantly higher on the items they selected than

on the nonselected items. A possible interpretation of such results may be that the fairness of evaluation might be increased considerably when items to be rated are consistent with teachers' instructional objectives and philosophies.

Often, instruments developed on the basis of common rating items do not take into consideration the differential teaching styles of instructors being evaluated. Research evidence (Gage, 1972) consistently shows that even the most competent teachers have their unique strengths and weaknesses, that no two teachers have the same teaching styles or objectives for any given course. Therefore, the most important problem to which the hypotheses of this study were addressed concerned the self-selection of relevant items by faculty. In consideration of such an idiographic procedure this study attempted to develop a systematic rating device to be used at the college level. The results of the study demonstrate that, when faculty are given an opportunity to select rating items which are compatible with their instructional objectives, student ratings can provide reasonably reliable and valid information concerning the quality of their educational experience.

The results of this investigation shed some light on the problem of using student ratings for determining instructional competence. Analysis of the data indicated that the instrument had desirable psychometric properties. Using Cronbach's Alpha (1951), the reliability of the 27 items (9-35) describing effective teaching was found to be .85. Other item intercorrelations showed that the instrument has high

internal consistency. Several validity data further indicated that items included in the instrument could be validly used for assessing diversified instructional approaches. Reliability and validity data obtained with the instrument also suggested the appropriateness of student ratings in spite of concern raised by several researchers (Kerlinger, 1971; Kossoff, 1972) in the field.

Implications

This study was primarily concerned with the development and validation of an instrument to measure teaching effectiveness. However, one ancillary concern was with the development of a valid body of knowledge related to effective teaching so that teacher education programs can be improved and upgraded, since few colleges in America are doing a first rate job of training students to become effective college or university teachers (Eble, 1971).

What is needed to upgrade and improve our teacher education programs? First, each institution needs to examine its fundamental goals, purposes, objectives, and its philosophy and principles of teacher evaluation (Stake, 1972). As an example, in recent years on many campuses much of instructional practices, testing, and evaluation have focused almost exclusively with cognitive information geared to factual recall in a form consistent with performance criteria or behavioral objectives (Combs et al., 1974). While no one questions the appropriateness of learning pertinent facts, factual information is not valuable as an end to itself. Such cognitively oriented information is only valuable as it increases human understanding and contributes to

effectiveness in living. Something more than an objective assessment of factual content material is needed. The goals of education involve affective as well as cognitive changes (Bloom et al., 1964). Thus a professional teacher education program necessitates a continuous evaluation procedure using student perceptions (ratings) of teacher influence. It is believed that student ratings reflect complex attitudinal learnings associated with the affective domain (Aleamoni and Spencer, 1971).

Second, since teaching, learning, and evaluation are integrally related continuous processes, teacher education programs must address themselves to how well course offerings and instructional activities make a constructive contribution to the lives of the learners (Hildebrand et al., 1971). Nothing else is as a basic a criterion of effective teaching as the impact a teacher makes in the lives of the learners. Wilhelms (1967) suggested that this is where the traditional system of education is weakest and falls short. Student perceptions provide relevant information concerning instructional practices and curricular offerings and, thus, help perpetuate institutional excellence. Apparently, for a large segment of American higher education this might suggest that institutions will have to reconsider their fundamental goals and philosophies if quality teaching and learning are to be encouraged and rewarded.

The most basic assumption of this study is the intriguing idea that feedback is one of the most precious things in life. It is believed that student feedback provides valuable information which might

help a prospective instructor or an experienced instructor change his teaching strategies and activities. Consequently, the heart of a teacher evaluation process lies in its diagnostic contribution for improving instruction. A systematic feedback program can help instructors become more aware of their own teaching activities and more sensitive to the effect of their behavior on students. Student ratings provide teachers with a feedback mechanism in which desirable teaching procedures and classroom interactions are more clearly perceived and reinforced.

Several obstacles tend to impede effective college teaching. One hindrance to effective teaching at any level is that so many things occur in the typical classroom that the untrained teacher is unaware of much that happens. This problem can be partly solved through learning (Hilgard and Bower, 1966). With practice and systematic feedback from students, teachers can become more aware of their classroom behavior.

Useful feedback focuses on specific ways to improve teaching behavior. A systematic evaluation form for rating teachers by students needs to contain items which provide concrete information that can lead to specific behavioral changes. Evaluation forms which contain vague generalizations (i.e., "the course content was good," "more classes should be taught this way," "the teacher had self-confidence") are not very helpful for improving teaching. On the other hand, items which provide specific behavioral referents (i.e., "too many topics were presented in this course," "more classes should be taught this way in which students become actively involved in class discussions," "the teacher showed self-confidence in demonstrating the significance and importance

of the subject matter") tend to be helpful. Thus, a major contribution to the improvement of instruction is the use of an evaluation form containing specific, concrete items which have the potential for modifying instructional practices.

Another impediment toward generating effective teaching stems from the fact that there is usually no systematic procedure for providing teachers with information concerning what they do in classrooms. Many administrators believe that their supervisory ratings are accurate measures of effective teaching. Teachers, however, often feel that supervisory ratings are highly unreliable and invalid (Tuckman and Oliver, 1968). Teachers usually perceive administrative evaluation as inadequate and threatening because of frequent disagreement between teachers and supervisors over appropriate objectives to be taught (McNeil, 1971). Teachers, however, are most likely to alter their teaching activities and methodologies when they are provided with information showing a discrepancy between what they want to do and what they are doing (McKeachie, 1969). Therefore, it is recommended that a systematic rating form be used in which instructors select relevant items.

It has been well documented (Hildebrand, 1972) that teachers rarely receive direct, systematic feedback from students or administrators concerning their instructional activities. Good and Brophy (1973) mentioned that the most substantial problem in improving teaching may be perceptual in nature--that the key to helping teachers change their behavior may be in providing feedback so that they can see themselves as well as students in action.

In consideration of this view, it would seem that video-tapes are the best approach for improving teaching activity. However, video feedback tends to be valuable only in the presence of a skilled consultant who provides directive, encouraging comments (Baker, 1970). A simpler, less expensive approach that seems just as efficient is the use of student ratings. When student ratings are used within the context of each instructor's objectives and frame of reference, this source of data appears to provide a viable feedback channel for promoting teacher growth (Gage, 1972; Centra, 1973).

In order to gain confidence in their teaching abilities, teachers need to receive progress toward their identified goals and objectives. When student ratings are used for formative (teacher improvement) evaluative purposes, involvement in a diagnostic teaching program presents opportunities for instructors to become aware of their increased teaching competence and enhanced self-esteem. Self-concept studies (Coopersmith, 1967) documented the importance of people feeling competent in learning new skills.

It is incumbent upon educators that they assume responsibility for assuring that teacher evaluation is an aid rather than a hindrance to faculty learning. Metcalfe (1972, p. 36) recommended that a teacher evaluation system should be administered so that each instructor will use the evaluation form in a manner to be most beneficial to that teacher: "Any evaluation system will be of absolutely no help to the professor unless he himself wants to improve. If he participates only because of pressure, and if he has a closed mind, evaluation will be of little or no help to him." The climate of formal evaluations is often

so threatening to block out honest self-appraisal. Especially, the less able or new teacher may feel that the main thrust of the program is geared to reveal as many as possible of his weaknesses and shortcomings. Self-concept research (Kelley, 1948; Purkey, 1970) documents that human beings are so constituted that they can look at themselves with open, nondefensive eyes only when they are in a relaxed, supportive, non-threatening situation. When teachers feel threatened, they tend to distort the feedback offered them to make it match their self-image. Consequently, traditional evaluation programs may lead more often to defensive self-distortion than to perceptive, honest self-appraisal. In such a threatening climate, teachers inadvertently forfeit their opportunity to learn about their real strengths and resources as well as one's weaknesses by being defensive and closed to feedback. Therefore, it is recommended that, if evaluation data are to be used for both formative and summative purposes, an instructor being evaluated not be held accountable for the first semester or two in which he teaches a course. He needs time and experience to polish his teaching skills and develop confidence in teaching this new subject matter domain. Thus, considerable care must be exercised so that any evaluation system does not acquire a punitive connotation or function.

Some researchers (Eble, 1971; Hildebrand et al., 1971) believe that the results of student ratings of instructors should be made available for the public record in the form of a published booklet. While this procedure may seem advantageous to students as an aid for course selection, it typically fails to consider each instructor's objectives

and unique approach to teaching. In this regard, overzealous published accounts of student ratings about faculty may present an unfair challenge or insensitive threat to improve one's teaching abilities. At this research stage of development teaching is too individualistic and complex an activity to assume that all instructors are alike and should be treated the same as if they were trained seals. Gage (1972) and Centra (1973), therefore, recommended that student ratings should be used mainly for instructional self-improvement in which the evaluative results are kept confidential.

Other researchers (Hildebrand, 1972; Miller, 1972) believe that student ratings should be used for making administrative decisions concerning teacher salaries and promotions. However, Gage (1972) and Centra (1973) recommended that teacher evaluations be used mainly for formative (growth) purposes in supplying diagnostic information for teacher improvement. They believed that self-appraisal is the most important function of student ratings. Centra (1973) mentioned that it would be helpful to have norms for teachers to interpret their ratings. Gage (1972) advocated that relative high and low ratings should be compared by the instructor with reference only to himself in order to provide valid, useful feedback. In this way, by maintaining the confidentiality of ratings, instructors will not be penalized by administrative actions. More research is needed in such areas in order to determine the impact of student ratings on administrative decisions.

Student ratings are especially valuable when looked at in the perspective of a two to three year cumulative period for each participating faculty member. Even though a highly effective instructor may

receive low ratings in one or more courses because of extraneous factors which are beyond his control (Gage, 1972), over an extended time period certain stable trends would likely emerge. Of course, these "stable" trends are subject to improvement and change, particularly if the faculty member desires to modify his teaching skills and procedures. This is where small seminar discussions might be especially effective in enhancing the instructor's teaching procedures and methods.

In addition to student feedback, efforts toward behavioral change in faculty might be encouraged by the social support, intellectual stimulation, and modeling from one's peers. Sanford (1968, p. 170) remarked, "College faculties seldom talk about what it's like in the classroom, what one is to do there, and how to deal with this situation or another." As one example, the Great Neck School District found that faculty peer evaluation resulted in significant behavioral changes in participants who observed one another's teaching (Kotcher and Doremus, 1972). Additionally, the results of ratings could be used to provide counseling services to instructors who received low ratings.

In light of the above-mentioned potentialities of student ratings, it is believed that the present work made contributions for promoting effective teaching and was specifically intended to facilitate the following. First, the student feedback was intended to stimulate faculty members to think more carefully about their course objectives and teaching procedures. Second, this systematic feedback was intended to instill motivation to improve teaching by providing comparative evaluation results of the ratings. In particular, it was assumed that

faculty members who received evaluations below their expectations would be motivated to improve their performance. In this respect this program was designed to provide diagnostic information needed for instructional improvement. Third, this feedback information could be used as an incentive device in the building of one's work record and for enhancing one's professional status. Therefore, at the discretion of each faculty member the results from the student ratings could be given to various administrators and committees for appointment, promotion, and tenure decisions. This way it was intended that each instructor would become responsible for improving the effectiveness of his teaching and for documenting his overall academic performance. Such a responsible, documentary procedure has been recommended by Bertramson (1972) and Sord (1973).

Limitations

Some instructors used this instrument in as many as two or three classes. The data of this study are based on 50 volunteer faculty members for 76 classes. Because some teachers used the instrument in more than one class, this fact may have reduced the variability of student responses and consequently lowered the reliability. In addition, some students may have filled out the instrument in more than one class and lost interest and motivation.

Because of the restricted, small sample (N=50) of participating faculty, the reliability (.85) of the instrument was not maximized. With a wider spectrum of sampling space (greater faculty participation)

the probability is greater that this obtained reliability value will be increased.

Additionally, the instrument developed in this study does not show cross validation data. A cross validation of the instrument needs to be conducted on other samples of students in different colleges and universities (Guilford, 1954).

During the instrument's administration, no time limit was imposed. However, informal feedback from students and faculty seemed to indicate that most students took 20-30 minutes to complete the entire evaluation form. Also, the 47-item final form was not administered at a uniform time during the last two weeks of the semester. Some instructors administered it two weeks before the final examination, some a day before the final examination, and some after the examination. We need to know the optimal time period to administer student rating forms that will be fair and equitable to all concerned.

Recommendations

The results of the present study indicated that the instrument thus developed could be used effectively for evaluating teaching effectiveness. Additionally, the results related to class size and faculty's option to select items from a measuring instrument are suggestive of certain sound educational practices. However, certain recommendations for future work in this area can be made. Such recommendations emerging from the results of the present study are summarized as follows:

(1) in spite of significant results of the study it is recommended that the study be replicated in various settings in order to enhance the

scope of the developed instrument; (2) the results of the study indicated that students rate the instructors significantly higher on the items selected by faculty than on the general type of item used on standard rating forms. Further studies are suggested to compare the effect of feedback for teachers who selected the items for their evaluation with those who used a standard form; (3) if the results of student ratings are to be used as feedback for the instructors, a number of studies should be directed toward the development of specific feedback models in order to facilitate the development of useful teaching skills in instructors; (4) it would seem desirable to widen the use and scope of the instrument by investigating the relationship between the ratings of an instructor and the measured outcomes of his students' learning; (5) it also can be recommended that the use of student ratings be made an integral part of a teaching-learning process. This means that evaluative data be used in developing specific services which could be available to the instructors who intend to upgrade their teaching skills. It is hoped that all teaching institutions in the future will have such a teaching-learning resource center; and (6) since the developed instrument reflects specific content domains (knowledge, affective, and behavioral), the results obtained from rating instruments can be meaningfully implemented in the development of course content and curricular materials for training educational change agents.

In summary, it is recommended that at the outset of a teacher evaluation program that faculty members be informed concerning the purposes, philosophies, and procedures attendant to such a systematic

endeavor. They need to be told honestly now student ratings will be used, whether for formative (growth) and/or summative (summary) evaluation purposes.

Even though any instrument for measuring teaching effectiveness is fallible with inherent limitations and shortcomings, such a rating instrument to be used for teacher self-appraisal seems adequately justified in light of the empirical findings and rationale reported in this study. Self-awareness then is a starting point for building a sense of accountability in one's teaching.

APPENDIX A

69-ITEM INITIAL INSTRUMENT

College of Education
University of Arizona
1972-73

Student Feedback on the Effectiveness of Instruction and Course

DIRECTIONS: Print your instructor's name and course number (or his code number) on the top of the answer sheet. Your name is to remain anonymous.

By giving thoughtful and honest answers to these questions, you will help your instructor improve this course and his teaching procedures. Careless or dishonest answers may have the opposite effect. Please answer all questions even though, in some instances, none of the alternatives expresses your reaction exactly.

Record all answers on the provided IBM answer sheet by blackening the appropriate space with a pencil (no. 2). Be sure to erase corrections carefully.

When you have finished you are encouraged to use the remaining space on the last page to make any comments which you feel might improve the course or help the instructor.

Part I. Describe the teaching procedures used by your instructor by using the following coded words with their relative time percentages:

- | | | | |
|---|---|--|---|
| 1 = Hardly Ever
(0 - 20%)
of the time | 2 = Occasionally
(21 - 40%)
of the time | 3 = Sometimes
(41 - 60%)
of the time | 4 = Frequently
(61 - 80%)
of the time |
| 5 = Almost Always
(81 - 100%)
of the time | | | |

The instructor:

1. Was enthusiastic about teaching the subject.
2. Answered student questions as completely as possible.
3. Presented sufficient examples or illustrations to help clarify the material.
4. Clearly stated the objectives of the course.
5. Summarized or stressed major points which aided learning and retention.
6. Stimulated me to put forth my best effort in learning all I could about the course and related material.
7. Discussed points of view other than his own.
8. Wasted class time.
9. Was readily available for individual help or consultation with students.

The instructor:

10. Accomplished his stated objectives for the course.
11. Clearly explained the course material.
12. Provided feedback within a reasonable time period on examination results or other academic work.
13. Encouraged students to ask questions and express their ideas and opinions.
14. Seemed well informed in knowledge of his subject matter area.
15. Gave examinations which reflected the important aspects and objectives of the course.
16. Stimulated independent thinking, reasoning, and decision making.
17. Monopolized class discussion, giving students little or no opportunity to speak.
18. Seemed well prepared for each class session.
19. Was flexible, changing his approach to meet student needs or new situations.
20. Showed concern when students were confused or had trouble understanding the course material.
21. Covered an appropriate number of topics in that an adequate amount of material was presented.
22. Made a significant impact on my outlook, philosophy, or style of life.
23. Presented material in such a way that I became bored and lost interest in what he was saying.
24. Appeared genuinely interested in student learning and progress by being actively helpful.
25. Provided students with an opportunity to apply concepts to demonstrate understanding.
26. Gave helpful feedback in commenting on exams or papers.
27. Presented material in a humorous way.
28. Stated his course requirements and expectations by telling students how they would be evaluated.
29. Failed to identify or distinguish between significant and insignificant material.
30. Stimulated thinking by raising challenging issues, questions, or problems.
31. Provided students with an opportunity to participate actively in some decision making processes directly related to the course.
32. Made good use of teaching aids (chalkboards, projectors, or models).
33. Assigned grades on a fair, reasonable basis.
34. Increased my interest in the course material and other related areas.
35. Spoke with expressiveness and variety in tone of voice.
36. Gave examinations stressing unnecessary rote memorization.
37. Was sensitive to student feelings and problems.
38. Showed self-confidence by demonstrating the significance and importance of his subject matter.
39. Was a model of how professionals in his field acquire new knowledge.

40. Encouraged and praised students at appropriate times.
 41. Found ways to help students answer their own questions.
 42. Gave assignments which seemed appropriate and fair.
-

Part II. Blacken one response number for each of the following questions. If an "other" response category is used, please write your comment(s) on separate paper.

43. Class time was wasted most frequently by:
 1. Not Applicable (Class time was hardly ever wasted).
 2. Dwelling on insignificant, irrelevant material.
 3. Going over repetitious points already covered.
 4. Doing administrative chores (taking roll, handing out, or collecting papers).
 5. Allowing too much student talk.
 6. Other (comment on separate paper _____).
 44. Concerning your preparation and ability, the level of difficulty of this course was:
 1. very easy
 2. somewhat easy
 3. about right
 4. somewhat difficult
 5. very difficult
 45. The amount of work for this course in relation to similar courses of equal credit was:
 1. much lighter
 2. lighter
 3. about the same
 4. heavier
 5. much heavier
 46. In general, compared to other high school or college teachers which you have had, how effective has the instructor been in this course?
 1. one of the most effective (top 10%)
 2. more effective than most (top 30%)
 3. about average
 4. not as effective as most (lowest 30%)
 5. one of the least effective (lowest 10%)
 47. The pace or rate at which the instructor covered the material during the semester was:
 1. very slow
 2. somewhat slow
 3. just about right
 4. somewhat fast
 5. very fast
-

Part III. SELF-EVALUATION. Describe your personal attitudes and behavior with respect to this course, using the following code:

1 = Definitely False 2 = More False Than True 3 = Inbetween
 4 = More True Than False 5 = Definitely True

48. I learned important principles, generalizations, methods, or theories.
49. I gained a broader understanding and appreciation of intellectual-cultural activity (music, science, literature, or art).
50. I discovered or learned some implications of the course material for understanding my interests, talents, and values.
51. I exercised skills in expressing myself orally and/or writing.
52. I developed a sense of personal responsibility (self-reliance, self-discipline) in applying what I learned.
53. I developed or exercised my creative problem solving abilities, skills, and interests.

Part IV. Answer either one of the following questions (54 and 55), depending on which one is most applicable to you. For "other" response categories, please comment on separate paper, indicating which numbered question you answered.

54. I would recommend the instructor to other students because of:
 1. his easy grading system.
 2. his easy to follow lectures or presentations.
 3. his interesting, dynamic speaking style.
 4. the learning I experienced.
 5. Other (Please comment on separate paper _____).
55. I would NOT recommend the instructor to other students because:
 1. he spoke in a rambling, incoherent fashion (failed to communicate).
 2. he would discuss exclusively his viewpoints and interests.
 3. he was seldom available for individual help or consultation.
 4. there was great disagreement between the announced objectives of the course and what was actually taught.
 5. I learned very little.
 6. Other (Please comment on separate paper _____).

Part V. Please provide the requested information by filling in the most appropriate response category. For "other" response categories, please comment on separate paper, indicating which numbered question you answered.

56. What grade (University of Arizona system) do you expect to receive in this course?

1. 1 (A)	4. 4 (D)
2. 2 (B)	5. Either (pass, fail, no credit)
3. 3 (C)	

57. What is your class level?
1. Junior
 2. Senior
 3. Graduate (Masters)
 4. Graduate (Doctoral)
58. Sex:
1. Female
 2. Male
59. What is your cumulative grade-point average? (Use Univ. of Ariz. system: 1=A, 2=B, 3=C, 4=D, 5=E)
1. 1.0 - 1.49
 2. 1.5 - 1.99
 3. 2.0 - 2.49
 4. 2.5 - 2.99
 5. 3.0 - 5.0
60. How was class size in relationship to the method of instruction used?
1. Satisfactory
 2. Class was too large
 3. Class was too small
 4. Class size did not make any difference
61. Which one of the following is the most important reason why you selected the course?
1. Interesting subject
 2. Required course
 3. Thought I could make a good grade
 4. Friend(s) or faculty advisor's recommendation
 5. Teacher's excellent reputation
 6. Other (Please comment on separate paper _____).
62. Select one of the following which best describes this course for you?
1. Elective not required in any way
 2. College requirement but not part of my major or minor field
 3. Minor requirement or required elective outside major field
 4. Major requirement or elective within major field
 5. Other (Please comment on separate paper _____).

Part VI. Blacken one response number for each of the following questions (63-69). Use the following key:

- 1 = Not Applicable, don't know, or there were none
- 2 = Poor
- 3 = Fair
- 4 = Good
- 5 = Excellent

63. In general, I would rate the assigned readings as
64. In general, I would rate the quality of exams as

- 65. Overall, I would rate the quality of lectures or class presentations as.
- 66. I would rate the overall value of class discussions as.
- 67. I would rate the laboratories, workshops, or seminars as.
- 68. In general, I would rate the textbook(s) as
- 69. I would rate the overall value of the course as

Part VII. If the instructor provides additional items, use these possible numbers (70-75) and rate them accordingly.

Part VIII. OPEN COMMENTS AND SUGGESTIONS. You might indicate those specific aspects of the instructor and course you liked MOST as well as those you liked LEAST. Also, it is helpful if you would state how you feel the course and the way it was taught could be improved. (Use separate paper)

APPENDIX B

52-ITEM INTERMEDIATE INSTRUMENT

DIRECTIONS: On the following pages are fifty-two items which have been found by numerous research studies to be related to effective teaching. Using the provided 1-6 rating scale below, please rate each item to the degree to which it fulfills your concept and feeling about the ideal, effective instructor and course. (Use a soft lead pencil (no. 2) in marking the IBM answer sheet)

RATING SCALE DEFINED:

1 = item describing the instructor and course the opposite of ideal and effective.

6 = item describing the ideal, effective instructor and course.

(Unideal, Ineffective 1 - 2 - 3 - 4 - 5 - 6 (Ideal, Effective
instructor (neutral) instructor)

Please return to me the completed answer sheet by Thursday, November 30. (For students: return the sheet to the classroom teacher or proctor)

When the items are eventually used in student evaluation of instruction, the student will rate on a five point scale the extent to which each statement applies to his instructor's effectiveness in the course. At present, we need to find out what statements are relevant to good (effective) instruction for our needs. Therefore, the purpose of this pilot study is to develop a reasonably reliable and valid instrument based on effective teaching as perceived by both faculty and students in the College of Education. Once this instrument is developed, it will be available to each faculty member in the spring of 1973.

In addition, you are invited to submit to me similar, yet different, items which you feel describe activities of the ideal, effective teacher and course.

Thank you very much,

David Annadale

David Annadale
Project Director
Committee on the Evaluation of Teaching
College of Education, Room 201

The instructor:

1. Encourages and praises students at appropriate times.
2. Speaks with expressiveness and variety in tone of voice.
3. Answers student questions as completely as possible.
4. Encourages students to develop a sense of personal responsibility (self-reliance, self-discipline) in applying what they learn.
5. Stimulates students to gain a broader understanding and appreciation of intellectual-cultural activity (music, science, literature, or art).
6. Has an interesting style of presentation.
7. Presents sufficient examples or illustrations to help clarify the material.
8. Finds ways to help students answer their own questions.
9. Is sensitive to student feelings and problems.
10. Teaches the course in such a way that students recommend him to other people.
11. Is enthusiastic about teaching the subject.
12. Encourages students to develop or exercise their own creative problem solving abilities, skills, and interests.
13. Clearly states the objectives of the course.
14. Assigns grades on a fair, reasonable basis.
15. Is one of the most effective teachers, compared to other instructors in high school or college.
16. Makes good use of teaching aids (chalkboards, projectors, or models).
17. Shows self-confidence by demonstrating the significance and importance of his subject matter.
18. Provides an opportunity for students to exercise skill in expressing themselves orally and/or writing.
19. Stresses important principles, generalizations, methods, or theories.
20. Clearly explains the course material.
21. Stimulates thinking by raising challenging issues, questions, or problems.
22. Uses class time effectively.
23. Seems well prepared for each class session.
24. Encourages students to discover some implications of the course material for understanding oneself (interests, talents, or values).

The instructor:

25. Provides students with an opportunity to participate actively in some decision making processes directly related to the course.
26. Discusses points of view other than his own.
27. Stimulates students to put forth their best effort in learning all they can about the course material and other related areas.
28. Summarizes or stresses major points which aid learning and retention.
29. Is readily available for individual help or consultation with students.
30. Accomplishes his stated objectives for the course.
31. Is flexible, changing his approach to meet student needs or new situations.
32. Shows concern when students are confused or have trouble understanding the course material.
33. Encourages students to ask questions and express their ideas or opinions.
34. Stimulates independent thinking, reasoning, and decision making.
35. Provides feedback within a reasonable time period on examination results or other academic work.
36. Gives examinations which reflect the important aspects and objectives of the course.
37. Seems well informed in knowledge of his subject matter area.
38. Covers an appropriate number of topics by presenting an adequate amount of material.
39. Makes a significant impact on student outlook, philosophy, or lifestyle.
40. Gives assignments which seem appropriate and fair.
41. States his course requirements and expectations by telling students how they will be evaluated.
42. Sometimes presents material in a humorous way.
43. Gives helpful feedback in commenting on exams or papers.
44. Paces the course in such a way that ideas and concepts are developed at an adequate rate.
45. Provides students with an opportunity to apply concepts to demonstrate understanding.
46. Appears genuinely interested in student learning and progress by being actively helpful.

The instructor:

47. Teaches the course at an appropriate level of difficulty in consideration of student ability and background preparation.
48. Gives conceptual examinations, stressing generalizations, principles, or trends.
49. Increases student interest in the course material and in other related areas.
50. Distinguishes between significant and insignificant material.
51. Requires an appropriate amount of work in relation to similar courses of equal credit.
52. Is a model of how professionals in his field acquire new knowledge.

APPENDIX C

47-ITEM FINAL FORM

College of Education
1973

Student Feedback on the Effectiveness of Instruction and Course

DIRECTIONS: Print your instructor's code number on the top of the answer sheet in the column marked MATRIC NO. Then blacken the appropriate spaces across from your instructor's code number. Your name is to remain anonymous, please.

By giving thoughtful and honest answers to these questions, you will help your instructor improve this course and his teaching procedures. Careless or dishonest answers may have the opposite effect. Please answer all questions even though, in some instances, none of the alternatives expresses your reaction exactly or, in some cases, the item is not applicable.

Record all answers on the provided IBM answer sheet by blackening the appropriate slot with a no. 2 pencil. Be sure to erase corrections carefully.

When you have finished you are encouraged to use the remaining space on the last page to make any comments which you feel might improve the course or help the instructor.

Part I. GENERAL INFORMATION ABOUT COURSE AND STUDENT. Please provide the requested information by filling in the most appropriate response category.

1. What time of the day does this class meet?
 1. 8 - 10 a.m.
 2. 10 a.m. - 12 p.m.
 3. 12 - 2 p.m.
 4. 2 - 4 p.m.
 5. 4 p.m. on
2. What grade (University of Arizona system) do you expect to receive in this course?
 1. 1 (A)
 2. 2 (B)
 3. 3 (C)
 4. 4 (D)
 5. Either (pass, fail, no credit)
3. The number of students in this class is approximately:
 1. 1-15
 2. 16-30
 3. 31-45
 4. 46-60 or higher

4. What is your class level?
 1. Junior
 2. Senior
 3. Graduate (Masters)
 4. Graduate (Doctoral)
 5. Other
5. Sex:
 1. Female
 2. Male
6. What is your cumulative grade-point average? (Use Univ. of Ariz. System: 1=A, 2=B, 3=C, 4=D, 5=F)
 1. 1.0 - 1.49
 2. 1.5 - 1.99
 3. 2.0 - 2.49
 4. 2.5 - 2.99
 5. 3.0 - 5.0
7. How was class size in relationship to the method of instruction used?
 1. Satisfactory
 2. Class was too large
 3. Class was too small
 4. Class size did not make any difference
8. Select one of the following which best describes this course for you.
 1. Elective not required in any way
 2. Elective within major field
 3. Elective within minor field
 4. Major requirement
 5. Minor requirement

Part II. SELF-EVALUATION. Describe your personal attitudes and behavior with respect to this course, using the following code:

1 = Definitely False 2 = More False Than True 3 = Inbetween
 4 = More True Than False 5 = Definitely True

9. I developed or learned specific skills, competencies, and points of view needed by professionals in the field most closely related to this course.
 10. I developed or exercised my creative problem solving abilities, skills, and interests.
 11. I discovered or learned some implications of the course material for understanding my interests, talents, and values.
 12. I developed a sense of personal responsibility (self-reliance, self-discipline) in applying what I learned.
 13. I exercised skills in expressing myself orally and/or writing.
-

Part III. TEACHER EVALUATION. Describe the teaching procedures used by your instructor by using the following coded words with their relative time percentages:

1 = Hardly Ever (0-20%) of the time	2 = Occasionally (21-40%) of the time	3 = Sometimes (41-60%) of the time
4 = Frequently (61-80%) of the time	5 = Almost Always (81-100%) of the time	

EXAMPLE: If you perceived that your instructor was enthusiastic about 30% of the time in your class, then the following item, "The instructor was enthusiastic in presenting the course material," would be marked 2 = Occasionally (21-40%) of the time in the appropriate place in the answer sheet: 1 = 2 = 3 = 4 = 5 = .

The Instructor:

14. Was willing to discuss points of view other than his own.
15. Clearly explained the course material.
16. Stimulated me to put forth my best effort in learning all I could about the course and related material.
17. Summarized or stressed major points which aided learning and retention.
18. Gave helpful feedback in commenting on exams or papers.
19. Wasted class time.
20. Provided students with an opportunity to apply concepts to demonstrate understanding.
21. Was well prepared for each class session.
22. Was flexible, willing to change his approach to meet student needs or new situations.
23. Accomplished his objectives for the course.
24. Monopolized class discussion, giving students little or no opportunity to speak.
25. Provided feedback within a reasonable time period on examination results or other academic work.
26. Found ways to help students answer their own questions.
27. Presented material in such a way that I became bored and lost interest in what he was saying.
28. Gave examinations or other academic work which were consistent with the important aspects and objectives of the course.
29. Encouraged and praised students at appropriate times.
30. Provided students with an opportunity to participate actively in some decision making processes directly related to the course.
31. Gave examinations or other academic work stressing unnecessary rote memorization.
32. Assigned grades on a fair, reasonable basis.
33. Stated his course requirements and expectations by telling students how they would be evaluated.

34. Stimulated thinking by raising challenging issues, questions, or problems.
35. Clearly conveyed the objectives of the course.

Part IV. INFORMATION ABOUT INSTRUCTOR AND COURSE. Blacken one response for each of the following questions.

36. Class time was wasted most frequently by:
 1. not applicable (class time was hardly ever wasted.)
 2. dwelling on insignificant, irrelevant material.
 3. going over repetitious points already covered.
 4. doing administrative chores (taking role, handing out, or collecting papers).
 5. allowing too much student talk.
 6. other (please comment at the end of part VII)
 37. Concerning your preparation and ability, the level of difficulty of this course was:
 1. very easy.
 2. somewhat easy..
 3. about right.
 4. somewhat difficult.
 5. very difficult.
 38. The amount of work for this course in relation to similar courses of equal credit was:
 1. much lighter.
 2. lighter.
 3. about the same.
 4. heavier.
 5. much heavier.
 39. The pace or rate at which the instructor covered the material during the semester was:
 1. very slow.
 2. somewhat slow.
 3. just about right.
 4. somewhat fast.
 5. very fast.
 40. (1 or 2) Answer EITHER part of this item, depending on which one is the most applicable to you. (Use part VII at the end for any additional comments which you might have about the instructor and course.)
 40. I would recommend the instructor to other students because of:
 1. the learning I experienced.
 - OR
 40. I would NOT recommend the instructor to other students because:
 2. I learned very little.
-

Part V. OVERALL RATINGS. Blacken one response number for each of the following questions (41-47). Use the following key:

- 1 = Not applicable, don't know, or there were none.
- 2 = Poor
- 3 = Fair
- 4 = Good
- 5 = Excellent

- 41. In general, I would rate the assigned readings as
- 42. In general, I would rate the quality of exams as
- 43. Overall, I would rate the quality of lectures or class presentations as
- 44. I would rate the overall value of class discussions as
- 45. I would rate the laboratories, workshops, or seminars as
- 46. In general, I would rate the textbook(s) as
- 47. I would rate the overall value of the course as

Part VI. ADDITIONAL ITEMS. If the instructor provides additional items, use these possible numbers (48-55) and rate them accordingly.

Part VII. OPEN COMMENTS AND SUGGESTIONS. Using the back of this paper, you might indicate those specific aspects of the instructor and course you liked MOST as well as those you liked LEAST. Also, it is helpful if you would state how you feel the course and the way it was taught could be improved. (Also, please comment on how you feel this instrument could be improved.)

APPENDIX D

LETTER TO FACULTY REQUESTING VOLUNTEERS
FOR THE TEACHER EVALUATION

April 24, 1973

Dear Faculty Member:

During the last two weeks of this semester, if you would like, your class(es) will have an opportunity to give you systematic feedback about the quality of your teaching. Last semester 82 classes provided faculty members with this kind of valuable information. This semester we are looking forward to even greater faculty participation, using the enclosed rating form which has been revised considerably. Now it should take only 15-20 minutes for students to complete the entire questionnaire.


Some possible uses and developmental goals of this type of evaluation instrument are presented on the following two pages. This is a tentative, guiding rationale and in no way speaks for the entire faculty or for all of the members of the Committee on the Evaluation of Teaching in the College of Education. Since this instrument has been developing and evolving all of this past academic year (1972-73), I anticipate that further changes, refinements, and elaborations will continue to be made in the future.

At this research stage in the instrument's development, an effort is being made to assure that each faculty member's unique style of teaching and objectives are carefully considered. This semester, using the enclosed rating form, I would like each faculty member who is participating to select those items which are compatible with his (her) own teaching goals, objectives, and unique style of teaching. In this way we can find out further information about two crucial points: (1) the individual differences and needs of the faculty and (2) the limitations and restrictions needed to be placed on the present instrument. In order to do this we need as much faculty involvement as possible.

If you would like to have your class(es) use copies of the enclosed evaluation form, please indicate below and return this paper to me at your earliest convenience.

Once again thank you for your cooperation.

Sincerely,



Dave Annadale, Project Director
Teaching Evaluation Committee
Room 201

(Please Print)

Name _____
No. of rating instruments needed: _____
Date needed _____

SOME POSSIBLE USES OF A TEACHING EVALUATION INSTRUMENT

1. Feedback for Professional Growth and Development. Self-appraisal of one's relative strengths (high ratings) and weaknesses (low ratings) can be compared for self-evaluation purposes.
2. A Source of Comprehensive Information. Teachers often have difficulty obtaining systematic information from informal discussions, tests, homework, and other assignments. A parallel can be drawn here to the automobile industry. It is one thing to give informal feedback to the car dealer or salesman. But the car industry knows that they can improve their product by requesting customers to fill out a comprehensive questionnaire. In this way systematic information can be used in improving the cars' performance. The same idea and rationale can be applied to student ratings of instruction.
3. Self-protection - from arbitrary dismissal from the university. An example is illustrated by several students who might not appreciate one's unique style of teaching and may complain to various administrators. The students who appreciate one's teaching do not usually drop by the department chairman or dean's office. This type of instrument would give all of your students an opportunity to express their perceptions and feelings about the quality of your teaching.
4. Building One's Work Record - in submitting one's annual report as evidence for promotion, retention, and tenure decisions. The text of the amended Chapter 8 (Academic Personnel Policy) as adopted by the Faculty Senate of the University of Arizona on November 6, 1971, states the following:

"8.01 Faculty members of the University of Arizona are appointed, retained or dismissed, and promoted on the basis of professional competence, teaching ability, scholarly attainment, and University and public service."
5. Confidentiality. Each instructor will be given the opportunity to select any random five digit number and use it as a reference source for identifying his results. In this way the positive growth aspects of the evaluation process will be emphasized.
6. Understanding its Limitations. In building norms, reliability, and validity of the instrument, it would be helpful to have as much faculty participation as possible. The more faculty involvement we have the more we can find out how the instrument is limited. Then the necessary revisions can be made to meet the individual instructional needs of the faculty.

Some Possible Developmental Goals of the Instrument

1. Correlating faculty and student perceptions and goals. An effort will be made to correlate both faculty and student expectations of

what constitutes effective teaching. These correlated items will be carefully considered in the validation of the instrument.

2. Selecting those items which are compatible with each faculty member's own unique style of teaching and objectives. An opportunity will be provided for faculty members to select their own objectives (items) based on research associated with effective teaching.
3. Correlating student ratings with learning. An attempt will be made to assess the extent to which student ratings correlate with measured learning outcomes.

APPENDIX E

LETTER TO FACULTY VOLUNTEERS FOR ADMINISTERING
THE 47-ITEM INSTRUMENT

April 24, 1973

Dear Dr.

I am glad that you volunteered to use the student rating form being developed by the Committee on the Evaluation of Teaching in the College of Education.

The present form has been shortened considerably and should take approximately 15-20 minutes to be completed.

In an attempt to assure objectivity, fairness, and confidentiality in my summary write-up of the analysis of the data of the entire evaluation results, I am requested that you use a code number instead of your name on the answer sheets. Please select any RANDOM five digit number (Example: 78932) and write it on the board in your class. Then tell your students to write it on the top middle column of the IBM answer sheet down from the MATRIC NO. which has an arrow pointing downward:

$$\begin{array}{l} \overline{7} \ 0 = 1 = 2 = 3 = 4 = 5 = 6 = 7 \blacktriangleright 8 = 9= \\ \overline{8} \ 0 = 1 = 2 = 3 = 4 = 5 = 6 = 7 = 8 \blacktriangleright 9= \\ \overline{9} \ 0 = 1 = 2 = 3 = 4 = 5 = 6 = 7 = 8 = 9\blacktriangleright \\ \overline{3} \ 0 = 1 = 2 = 3 \blacktriangleright 4 = 5 = 6 = 7 = 8 = 9= \\ \overline{2} \ 0 = 1 = 2 \blacktriangleright 3 = 4 = 5 = 6 = 7 = 8 = 9= \\ \overline{\quad} \ 0 = 1 = 2 = 3 = 4 = 5 = 6 = 7 = 8 = 9= \end{array}$$

Then ask your students to follow the directions on the provided rating form. After you have collected the rating forms and answer sheets, please make sure that every answer sheet has your code number blackened in the appropriate spaces across from the MATRIC NO. as illustrated above. Otherwise, the computer will not be able to score the returns. (SPECIAL NOTE: be sure to write down your code number some place so that you will not inadvertently forget it or confuse your number with someone else. This code number is your only way to identify your results.)

In case you have one to eight additional items or objectives for students to rate, use Part VII (items 48-55) for this aspect of the evaluation. For any additional items which you might have please write them on the board and indicate to your students how (on what basis or standard) you want them to rate those items. (If applicable, please furnish me with a copy of any additional items used plus a keyed answer sheet).

When the students complete the evaluation, return the IBM answer sheets and the entire evaluation forms in a large sealed envelope to your department secretary. She will then return the envelope to me and identify the department. In this way, we can receive further data about departments. When the computer print out and the summary write-up are available about three - four weeks later, your departmental secretary will be notified so she can pick up the report from my office

and return it to you. All of the students' open-ended comments will also be returned to you at that time.

In order to use a standard procedure, I am recommending that you administer the form to your own class(es). Also, I am recommending that the items be rated on the basis of the printed directions on the form. Individual questions can be handled with minimal deviations whenever necessary. As an example, if a student asks, "What does this mean by "stimulated" or "flexible," etc., you could respond as follows: "Whatever the word means to you," and not comment any further, if at all possible.

Each department secretary will be furnished with a supply pool of No. 2 pencils to be used by the faculty. Please return the pencils to the secretary so your colleagues can use them for their classes. In the event that you need extra pencils, the secretary in Dean Paulsen's office in room 201 will have a few extra ones in case you run short.

I would appreciate it if you would return all left over materials such as IBM answer sheets to me. Also, if you have any questions or suggestions for improving the instrument, I would like to hear from you.

Thank you for your cooperation and assistance.

Sincerely,



Dave Annadale
Project Director
Committee on the Evaluation of
Teaching, Room 201

APPENDIX F

LETTER TO FACULTY VOLUNTEERS ASKING FOR
THE SELECTION OF RELEVANT ITEMS

April 24, 1973

MEMORANDUM TO: Faculty Members Using the Student Evaluation Form

Regarding: SELECTION OF ONE'S OWN RELEVANT ITEMS

Dear Faculty Member:

I would like to suggest an interesting idea or procedure which I feel might lead to some significant professional growth and development in one's teaching. This procedure also will help to understand better the function and usefulness of this rating form.

After you pass out the forms to the students and have answered any questions, tell them you are going to leave the room for about 15-20 minutes. If they have any further questions which might come up while you are out of the room, tell them you will try to answer them when you return. (At some leading universities it is common protocol for professors not to be present in the room during the student evaluation. This leads to greater accuracy and truthfulness in the quality of the evaluation). Once you leave the room please identify those items on the rating form which provide valuable information and/or are compatible with your own objectives and style of teaching. By doing this, you will be provided with valuable feedback which is relevant to your own objectives and frame of reference.

More specifically, please CIRCLE the number beside those items on which you feel will provide valuable information and/or that you will be rated most fairly and effectively.

SPECIAL NOTE: Negatively worded items (i.e., "#19. Wasted Class time, #24. Monopolized class discussion, giving students little or no opportunity to speak, #27. Presented material in such a way that I became bored and lost interest in what he was saying, and #31. Gave examinations or other academic work stressing unnecessary rote memorization") are reversal items. These items help assure objectivity in student responding and will be scored and worded positively on the computer print out. Example. #19. Wasted class time really means the opposite: Used class time well. Please consider these reversal items in selecting your relevant objectives or items.

When you return the entire "package" of answer sheets and rating forms, please place the enclosed rating form with your circled, self-selected items on the top of the returned information.

Thank you very much.

Sincerely,



Dave Annadale, Project Director
Teaching Evaluation Committee

REFERENCES

- Ahmann, J. S. Role of evaluation in educational innovation. In Accountability in higher education in agriculture. Bozeman, Montana: National Association of State Universities and Land-Grant Colleges, 1972.
- Aleamoni, L. and Spencer, R. The Illinois course evaluation questionnaire: A description of its development and a report of some of its results. Research Report No. 92. Champaign, Illinois: Office of Instructional Resources, May, 1971.
- Amidon, J. and Hough, J. B. (Eds.) Interaction analysis: Theory, research, and application. Palo Alto, California: Addison-Wesley, 1967.
- Astin, A. and Lee, C. Current practices in the evaluation and training of college teachers. In C. B. Lee (Ed.) Improving College Teaching. Washington, D.C.: American Council on Education, 1967.
- Baker, H. Film and videotape feedback: A review of the literature. Report Series No. 53. Austin: Research and Development Center for Teacher Education, University of Texas, 1970.
- Bandura, A. Principles of behavior modification. New York: Holt, Rinehart and Winston, 1969.
- Bandura, A. and Walters, R. Social learning and personality development. New York: Holt, Rinehart and Winston, 1963.
- Bertramson, B. R. Evaluating teaching performance: How it is done. In Accountability in higher education in agriculture. Bozeman, Montana: National Association of State Universities and Land-Grant Colleges, 1972.
- Fiddle, B. J. The integration of teacher effectiveness research. In B. J. Fiddle and W. J. Ellena (Eds.) Contemporary research on teacher effectiveness. New York: Holt, Rinehart and Winston, 1964.
- Bloom, B. S., Krathwohl, D. R. and Masia, B. B. Taxonomy of educational objectives. Handbooks 1. Cognitive Domain; 2. Affective Domain. New York: David McKay, 1964.
- Bruner, J. S. Toward a theory of instruction. New York: W. W. Norton, 1966.

- Bryan, R. C. Reactions to teachers by students, parents, and administrators. Kalamazoo, Michigan: Western Michigan University, 1963.
- Centra, J. A. The student instructional report: Its development and uses. SIR Report No. 1. Princeton, N.J.: Educational Testing Service, 1972a.
- Centra, J. A. Two studies on the utility of student ratings for improving teaching. SIR Report No. 2. Princeton, N.J.: Educational Testing Service, 1972b.
- Centra, J. A. The student as godfather? The impact of student ratings on academia. Educational Researcher, 1973, 10, 4-8.
- Combs, A. W. Can we measure good teaching objectively? National Education Association Journal, 1964, 53, 34-36.
- Combs, A. W., Blume, R. A., Newman, A. J. and Hannelore, L. W. The professional education of teachers: A humanistic approach to teacher preparation. Boston: Allyn and Bacon, 1974.
- Coopersmith, S. The antecedents of self-esteem. San Francisco: W. H. Freeman, 1967.
- Costin, F., Greenough, W. and Menges, R. Student ratings of college teaching: Reliability, validity, and usefulness. Review of Educational Research, 1971, 5, 511-535.
- Cronbach, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.
- Downie, N. M. Student evaluation of faculty. Journal of Higher Education, 1952, 23, 495-496.
- Drucker, A. J. and Remmers, H. H. Do alumni and students differ in their attitudes toward instructors? Journal of Educational Psychology, 1951, 42, 129-143.
- Eble, K. E. The recognition and evaluation of teaching. Salt Lake City, Utah: Project to Improve College Teaching, 1971.
- Elliot, D. Characteristics and relationships of various criteria of college and university teaching. Purdue University Studies in Higher Education, 1950, No. 70, 5-61.
- Fishman, J. A. Cross-cultural perspective on the evaluation of guided behavioral change. In The evaluation of teaching. Washington, D.C.: Pi Lambda Theta, 1967.
- Flanagan, J. C. The critical incident technique. Psychological Bulletin, 1954, 51, 327-358.

- Flanders, N. A. Teacher influence, pupil attitudes, and achievement. Minneapolis: University of Minnesota Press, 1960.
- Flanders, N. A. Teacher influence in the classroom. In E. Amidon and J. Hough (Eds.). Interaction analysis: Theory, research, and application. Palo Alto, California: Addison-Wesley, 1967a.
- Flanders, N. A. Some relationships among teacher influence, pupil attitudes, and achievement. In E. Amidon and J. Hough (Eds.). Interaction analysis: Theory, research and application. Palo Alto, California: Addison-Wesley, 1967b.
- Frey, P. Comparative judgment scaling of student course ratings. American Educational Research Journal, 1973, 10, 149-154.
- Gage, N. L. The appraisal of college teaching. Journal of Higher Education, 1961, 32, 17-22.
- Gage, N. L. (Ed.). Handbook of research on teaching. Chicago: Rand McNally, 1963.
- Gage, N. L. Teacher effectiveness and teacher education: A search for a scientific basis. Palo Alto, California: Pacific Books, 1972.
- Gage, N. L., Runkel, P. J. and Chatterjee, B. B. Changing teacher behavior through feedback from pupils. In W. W. Charters and N. L. Gage (Eds.) Readings in the social psychology of education. Boston: Allyn and Bacon, 1963.
- Gagné, R. M. The conditions of learning. New York: Holt, Rinehart and Winston, 1965.
- Glaser, R. Instructional technology and the measurement of learning outcomes: Some questions. American Psychologist, 1963, 18, 519-521.
- Good, T. L. and Brophy, J. E. Looking in classrooms. New York: Harper and Row, 1973.
- Guilford, J. P. Psychometric methods. New York: McGraw-Hill, 1954.
- Gustad, J. Policies and practices in faculty evaluation. The Educational Record, 1961, 42, 192-211.
- Hamachek, D. E. What research tells us about the characteristics of "good" and "bad" teachers. In D. E. Hamachek (Ed.) Human dynamics in psychology and education. Boston: Allyn and Bacon, 1968.
- Hartley, E. and Hogan, T. Some additional factors in student evaluation of courses. American Educational Research Journal, 1972, 9, 241-250.

- Hess, C. V. Evaluating teaching performance--why should it be done? In Accountability in higher education in agriculture. Bozeman, Montana: National Association of State Universities and Land-Grant Colleges, 1972.
- Hildebrand, M. How to recommend promotion for a mediocre teacher without actually lying. Journal of Higher Education. January, 1972, 44-62.
- Hildebrand, M., Wilson, R., and Dienst, E. Evaluating university teaching. Berkeley, California: Center for Research and Development in Higher Education, 1971.
- Hilgard, E. R. and Bower, G. H. Theories of learning. New York: Appleton-Century-Crofts, 1966.
- Horst, P. Psychological measurement and prediction. Belmont, California: Brooks/Cole, 1968.
- Hoyt, D. P. Instructional effectiveness: I. Measurement of effectiveness, II. Identifying effective classroom procedures, and III. Improving instruction through student feedback. Manhattan: Kansas State University, 1969.
- Hunt, D. and Joyce, B. Teacher trainee personality and initial teaching style. American Educational Research Journal, 1967, 3, 253-259.
- Jencks, C. and Riesman, D. The academic revolution. Garden City, N.Y.: Doubleday, 1968.
- Kelley, E. C. Education for what is real. New York: Harper, 1948.
- Kerlinger, F. N. Attitudes toward education and perception of teacher characteristics: A Q study. American Educational Research Association Journal, 1966, 3, 159-168.
- Kerlinger, F. N. Student evaluation of university professors. School and Society, 1971, 99, 353-356.
- King, M. W. Academic freedom and accountability. In Accountability in higher education in agriculture. Bozeman, Montana: National Association of State Universities and Land-Grant Colleges, 1972.
- Knoell, D. Prediction of teaching success from word fluency data. Journal of Educational Research, 1953, 46, 673-683.
- Kossoff, E. Evaluating college professors by "scientific" methods. The American Scholar, 1972, 41, 79-93.

- Kotcher, E. and Doremus, R. R. Increasing positive interactive classroom behavior. Great Neck, New York: Great Neck Public Schools, 1972.
- Langen, T. D. Student assessment of teaching effectiveness. Improving College and University Teaching, Winter, 1966, 14, 22-25.
- Lessinger, L. Accountability in higher education: Policy springboard for reform. Breezy Point, Minnesota: Paper presented at the Minnesota Intercollegiate Faculty Conference, May 7, 1971, 1-4.
- McDonald, F. and Allen, D. Training effects of feedback and modeling procedures on teaching performance. Palo Alto, California: Stanford University Press, 1967.
- McKeachie, W. J. Research on teaching at the college and university level. In N. L. Gage (Ed.). Handbook of research on teaching. Chicago: Rand McNally and Company, 1963.
- McKeachie, W. J. Research in teaching: The gap between theory and practice. In C. B. Lee (Ed.). Improving college teaching. Washington, D.C.: American Council on Education, 1967.
- McKeachie, W. Student ratings of faculty. AAUP Bulletin, 1969, 55, 439-444.
- McKeachie, W. J. Improving teaching effectiveness. Washington, D.C.: Office of Education, ERIC Report No. ED 054 083, 1970.
- McKeachie, W., Lin, Y. and Mann, W. Student ratings of teacher effectiveness: Validity studies. American Educational Research Journal, 1971, 8, 435-445.
- McNeil, J. Toward accountable teachers: Their appraisal and improvement. New York: Holt, Rinehart and Winston, 1971.
- Metcalf, D. S. Evaluating teaching performance: Why should it be done? In Accountability in higher education in agriculture. Bozeman, Montana: National Association of State Universities and Land-Grant Colleges, 1972.
- Miller, R. Evaluating faculty performance. San Francisco: Jossey-Bass, 1972.
- Mitzel, H. Teacher effectiveness. In C. Harris (Ed.) Encyclopedia of educational research. New York: Macmillan Company, 1960.
- Morsh, J. E., Burgess, G. G. and Smith, P. N. Student achievement as a measure of instructor effectiveness. Journal of Educational Psychology, 1956, 47, 79-88.

- Popham, J. Performance tests of teaching proficiency: Rationale, development, and validation. American Educational Research Journal, 1971, 1, 105-117.
- Popham, J., and Baker, E. Systematic instruction. Englewood Cliffs, N.J.: Prentice-Hall, 1970.
- Purkey, W. W. Self-concept and school achievement. Englewood Cliffs, N.J.: Prentice-Hall, 1970.
- Remmers, H. H. Manual, the Purdue rating scale for instruction. West Lafayette, Indiana: University Book Store, 1960.
- Remmers, H. H. Rating methods in research on teaching. In N. Gage (Ed.) Handbook of research on teaching. Chicago: Rand McNally, 1963.
- Rogers, C. R. Freedom to learn. Columbus, Ohio: Charles Merrill, 1969.
- Russell, H. E. and Bendig, A. W. An investigation of the relation of student ratings of psychology instructors to the course achievement when academic aptitude is controlled. Educational and Psychological Measurement, 1953, 13, 626-635.
- Ryans, D. Characteristics of teachers. Washington, D.C.: American Council on Education, 1960.
- Ryans, D. G. Teacher behavior can be evaluated. In The evaluation of teaching. Washington, D.C.: Pi Lambda Theta, 1967.
- Sanford, N. Where colleges fail: A study of the student as a person. San Francisco: Jossey-Bass, 1968.
- Sciara, F. J. and Jantz, R. K. Accountability in American education. Boston: Allyn and Bacon, 1972.
- Scranton, W. (Ed.). President's commission report on campus unrest. Washington, D.C.: Government Printing Office, 1970.
- Solomon, D., Rosenberg, L. and Bazdek, W. Teacher behavior and student learning. Journal of Educational Psychology, 1964, 59, 297-301.
- Sord, B. H. Teaching effectiveness. Phi Kappa Phi Journal, 1973, 1, 58-66.
- Spaulding, R. Achievement, creativity, and self-concept correlates of teacher-pupil transactions in elementary schools. Urbana, Ill.: University of Illinois, 1963.

- Stake, R. E. Philosophy and principles of evaluating teaching performance. In Accountability in higher education in agriculture. Bozeman, Montana: National Association of State Universities and Land-Grant Colleges, 1972.
- Tuckman, B. and Oliver, W. Effectiveness of feedback to teachers as a function of source. Journal of Educational Psychology, 1968, 59, 297-301.
- Weaver, W. Good teaching. Science, 1966, 151, 1335.
- Wilhelms, F. T. (Ed.) Evaluation as feedback and guide. Washington, D.C.: Association for Supervision and Curriculum Development, 1967.
- Wilson, R., Gaff, J. and Bavry, J. Manual of information: Faculty characteristics questionnaire. Berkeley, California: Center for Research and Development in Higher Education, 1970.