1991

A study of peer involvement in the formative evaluation of instruction in higher education

Larry W. Keig
University of Northern Iowa

Order Number 9131189

A study of peer involvement in the formative evaluation of instruction in higher education

Keig, Larry William, Ed.D.

University of Northern Iowa, 1991

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

A STUDY OF PEER INVOLVEMENT IN THE FORMATIVE EVALUATION OF INSTRUCTION IN HIGHER EDUCATION

A Dissertation

Submitted

In Partial Fulfillment

of the Requirements for the Degree of

Doctor of Education

Approved:

Dr. Dale R. Jackson

Dr. Thomas Hansmeier

Dr. Basil Reppas

Dr. Bruce Rogers

Dr. Michael Waggoner

Dr. B. Wylie Anderson

Larry W. Keig

University of Northern Iowa

May 1991

DEDICATION

To the memory of my mother,

Frances M. Windhurst Keig,

who by example showed me--and my sisters--

how inherently joyful lifelong learning is, and can be,

and to my father,

William H. Keig,

who has modeled the virtues of

autonomy, diligence, and persistence,

qualities which have been, and continue to be,

of inestimable value to the three of us.

ACKNOWLEDGEMENTS

At the time I began doctoral study, I was determined to look at the field of educational administration in as much depth as time and my abilities allowed and to relate this study to my experiences as a college teacher as well as earn the credential to which the program of study would lead. I feel fortunate to have had the time and means to achieve, at least in part, all of these ends. I am grateful to my professors and colleagues in graduate programs for helping me to achieve these goals.

Over a period of several months of intensive work on this study, I have become indebted to the members of my dissertation committee for their help. I am especially grateful to Dr. Dale Jackson, Chair, for his advice, assistance, encouragement, and enthusiasm. I will remember with fondness the "marathon" sessions in which much of our significant work was accomplished and in which our personal friendship developed.

I am also thankful for the unselfish help of Drs. Wylie Anderson, Tom Hansmeier, Basil Reppas, Bruce Rogers, and Mike Waggoner. Dr. Hansmeier has been an inspiring teacher and mentor. I appreciated his friendship, counsel, and cheerfulness, not to mention his willingness to edit early and sometimes very rough drafts of the manuscript. Dr. Rogers' expertise in research design and statistics contributed significantly to the completion of this project. I appreciated Dr. Rogers' compassion, friendliness, and patience.

Dr. Reppas has been an advisor, confidante, and sounding board throughout the period of doctoral study. I have appreciated the manner in which Dr. Reppas has challenged me to look at educational issues critically and analytically. Dr. Waggoner has been a constant source of ideas and encouragement. Some of my most pleasant moments at Northern Iowa have come in our serious and not-so-serious discussions. Although I have had less involvement with Dr. Anderson, as he is an economist rather than a faculty member in the College of Education, I have appreciated his sound advice and keen intellect.

Much of the technical work for the dissertation study was done at the Center for Social and Behavioral Research (CSBR) by or under the direction of Mr. Robert Kramer. Mr. Kramer's expertise in survey research was of immense value at many stages of this project. I appreciated Mr. Kramer's always cheerful advice and counsel.

In addition to faculty members serving on the dissertation committee, I am also grateful to current and former faculty members for advice along the way. They include Drs. Don Carver, Charles Dedrick, Sue Doody, Larry Kavich, John K. Smith, and Marlene Strathe. I have also enjoyed discussions with Drs. Bob Decker, Jack Kimball, Ed Rutkowski, and Don Schmits.

The staff at Donald O. Rod Library has been of great help. I am especially indebted to Georgia Quirk and Jim Shaw who devoted large blocks of time to help me. I also appreciate the assistance of Larry Kieffer and Ed Wagner.

I am most grateful to Mrs. Marlene Shea for her assistance. Mrs. Shea's word processing skills and knowledge of the style requirements of the Graduate College were a comfort with other pressing matters leading up to the completion of the study.

Finally, I want to express my thanks to my sisters, Kathy Kerber and Barb Chamberlain, and their husbands, John Kerber and Don Chamberlain. Their kindness over the years of study have made the work less burdensome than it might otherwise have been.

Larry Keig

April, 1991

TABLE OF CONTENTS

CHAPTER PAGE

CHAPTER PAGE

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

CHAPTER I

THE RESEARCH PROBLEM AND ITS CLARIFYING COMPONENTS

Arguments for and against the utility of formally evaluating teacher performance, the value of different methods of evaluation, and the suitability of various constituencies in judging the instructional effectiveness of the faculty in higher education have been debated for decades. Despite these discussions, faculty members have been (and continue to be) evaluated, formally or informally, by individuals or groups of persons. These judgments have affected faculty members' professional and personal relationships with students, administrators, and colleagues. Evaluation has also had an effect, of course, on decisions regarding the reappointment, promotion, tenure, and compensation of faculty members.

Cohen and McKeachie (1981) and Scriven (1980) indicated that it is in the best interest of the faculty to have a formal, systematic faculty evaluation program in place, as it can protect faculty members from unjust personnel decisions. Lee (1982), Smith (1981), and Stevens (1985) suggested that a program of summative evaluation--evaluation conducted for decisions dealing with reappointment, promotion, tenure, and compensation--is also in the best interest of colleges and universities in an era when litigation sometimes results from personnel decisions unfavorable to faculty members, in that it is a means of demonstrating that the process

has been governed by reason, rather than handled arbitrarily or capriciously.

Until recently, evaluation was in place primarily for making personnel decisions, and little real attention was paid, for the most part, to improving instruction. Evaluation specifically designed to improve performance, commonly called formative evaluation, began in the late 1960s and early 1970s on a number of college and university campuses.

Statement of the Problem

The purpose of this study was to examine faculty attitudes toward methods of peer review in the formative evaluation of instruction in higher education and toward selected variables that may affect the willingness of faculty members to participate in these methods of evaluation. The four methods investigated were direct classroom observation, videotaping of classes, evaluation of course materials, and evaluation of instructor-graded student assignments.

A random sample of the full-time faculty of the baccalaureate degree-granting member institutions affiliated with the Iowa Association of Independent Colleges and Universities was requested to respond to a questionnaire dealing with these methods and other variables. Tests of statistical significance, correlational procedures, and descriptive statistics--frequency counts, percentages, means, standard deviations, and medians--were used in the data analysis. It was expected that this study would show

that faculty members would avail themselves of formative evaluation if it were in place, and that willingness to do so and attitudes toward what are called detractors, enhancers, and individual/institutional improvement factors would be related.

Need for the Study

This study was undertaken because there is a discrepancy between the prevailing view that college teaching would be improved by implementing programs of formative evaluation involving colleagues as evaluators of instruction, and the belief that such programs are not a common component of faculty evaluation. Therefore, it was decided to investigate a number of factors that might affect the willingness of faculty members to participate in such programs, and to explore relationships among faculty members who would participate and their attitudes about the detractors, enhancers and individual/institutional improvement factors.

Conceptual Framework

Motivating faculty to improve the quality of classroom teaching has, no doubt, been positively affected to a degree by the rewards dispensed (or withheld) through summative evaluation. Yet, faculty members have complained that summative evaluation has not validly measured teaching performance. It was also suggested that extrinsic rewards (e.g., tenure, promotion, and salary increases) usually maintain, rather than improve, performance (Hodgkinson, 1972).

Blackburn and Clark (1975), Cross (1986), Dressel (1976), and Hodgkinson (1972) were convinced that formative evaluation is more

promising than summative evaluation for improving the quality of college teaching. These researchers have insisted that instructional improvement, and not the emphasis normally placed on the making of personnel decisions, should be the first priority in faculty evaluation.

Soderberg (1986) proposed a three-dimensional instructional evaluation/improvement model based on teaching and learning theories. Its first dimension was represented by three time phases in the instructional process: pre-interaction (the planning period prior to delivery of instruction), interaction (the period when instruction is actually delivered, or when students are otherwise engaged in learning), and revision (a post-instructional period of reflection on, and evaluation of, events occurring in the pre-interactive and interactive phases). Soderberg thought of this dimension as a continuous and circular process.

The second dimension was represented by the means chosen to accomplish the events planned for each phase of the time dimension. It included the statement of goals and objectives, the selection of methods and materials for delivering the goals and objectives, and the types of feedback chosen to measure the degree of success of the interaction between the goals and objectives and the methods and materials selected to accomplish them.

Information sources constituted the third dimension. Soderberg recommended that four sources, or constituencies, be involved in evaluating teaching effectiveness: students (one of two primary

sources of information), informed peers (the second primary source), informed academic administrators, and self-assessment.

Soderberg's aim was to create a model credible enough to be accepted by faculty members as a valid means for evaluating instructional effectiveness. If it is credible to them, Soderberg theorized that evaluation would facilitate teaching improvement as well as serve as a vehicle for making personnel decisions.

In a study of colleague involvement in classroom observation and, to a lesser degree, of other methods of summative and formative evaluation, Britt (1982) addressed several issues regarding the involvement of peers in instructional evaluation, and found general support for peer review. Britt's study and Soderberg's study were important sources in the conceptualization of the research questions and hypotheses investigated in this study.

The Clarifying Components

Definition of Terms

In order to clarify terminology used in this study, the following definitions were used:

Attitude. An attitude is a state of mind or feeling with regard to, or a disposition toward, some matter.

Detractor variables. Detractor variables are factors that have the potential to take away from the value of the methods of evaluation investigated in this study.

Enhancer variables. Enhancer variables are factors that have the potential to add or contribute to the effectiveness of the methods of evaluation investigated in this study.

Formative evaluation. Formative evaluation is for purposes of improvement. It is accomplished by self-identification of strengths and weaknesses through interaction with, and with feedback from, one or more individuals who are knowledgeable about teaching and learning strategies and who are familiar with the teaching performance of the instructor being evaluated. Findings from formative evaluation are not intended to be used in the making of personnel decisions (i.e., those concerning reappointment, promotion, tenure, and/or compensation).

Full-time faculty member. A full-time faculty member teaches nine or more hours (or the equivalent) of course work each semester, and does not hold an appointment as an academic administrator above the departmental (or divisional) level.

Individual/institutional improvement factors. The individual/institutional improvement factors are factors that may be of benefit to the college or university and/or to faculty members and students.

Informed peer. An informed peer is a colleague who is knowledgeable by training and/or experience about teaching and learning strategies and methods of evaluation, and who is familiar with the objectives, methods of instruction, and instructional materials of the colleague being evaluated. That person may hold

any academic rank, but may not have the official responsibility of evaluating instruction for purposes of reappointment, promotion, tenure, or compensation.

Iowa Association of Independent Colleges and Universities. The Iowa Association of Independent Colleges and Universities (hereafter referred to as the IAICU) is a consortium of 28 private postsecondary institutions in the state of Iowa. Twenty-six of them offer at least the bachelor's degree, one is a two-year college, and the other is a business school. The IAICU maintains offices in Des Moines.

Methods of faculty evaluation. Methods of faculty evaluation are the approaches used in ascertaining the strengths and weaknesses and/or merit and worth of faculty members. The methods investigated in this study were direct classroom observation, videotaping of classes, evaluation of course materials, and evaluation of instructor-graded student assignments.

Peer evaluation. Peer evaluation, or peer review, is an assessment conducted by members of the faculty (i.e., colleagues) as opposed to that conducted by academic administrators or students.

Summative evaluation. Summative evaluation is an assessment of competency primarily for the purpose of making personnel decisions (i.e., reappointment, promotion, tenure, and/or compensation). It is not intended expressly for the purpose of improving teaching performance.

Assumptions

The following assumptions were formulated after extensive reading of the literature and/or after an examination of results of a pilot study in which information about formative evaluation was solicited from academic officers of the IAICU (Keig, 1989).

1. Most faculty members have a desire to improve their teaching (Blackburn & Clark, 1975; Farmer, 1976).

2. Faculty members at most colleges and universities have limited formal opportunities to improve their teaching (Keig, 1989).

3. Systematic programs to help faculty members improve their teaching are not available at most colleges and universities. In other words, formative evaluation is not commonly employed at most institutions (Keig, 1989).

4. Findings from formative evaluation are usually shared through discussions between/among evaluator(s) and the faculty member being evaluated, and are not normally communicated in writing or numerically from rating scales (Brock, 1981).

Limitations of the Study

In order to restrict the scope of the study, the following limitations were imposed.

1. The study focused on formative evaluation, not on summative evaluation. Every item on the questionnaire, and each research question and hypothesis, was designed to examine formative evaluation.

2. The study was restricted to an investigation of teaching performance, not the evaluation of research and publication or service.

3. The study focused on the role that might be played by peers in the formative evaluation of teaching performance. The roles played by academic administrators and/or students, or self-assessment except as it is associated with peer review, were not investigated.

4. The questionnaire was sent only to faculty members at the 26 baccalaureate degree-granting member institutions of the IAICU. It was not sent to faculty members at the IAICU member institutions that do not offer the bachelor's degree or to faculty members at the three regents' institutions.

Design Components for the Study

Research Questions and Hypotheses

Four research questions were studied and hypotheses were formulated for each of these questions. These research questions and hypotheses were:

Research Question 1. What proportion of faculty members would avail themselves of direct classroom observation, videotaping of classes, evaluation of course materials, and evaluation of instructor-graded student assignments if these methods were available for the purpose of formative evaluation?

Hypothesis 1: A majority of faculty members would avail themselves of direct classroom observation, videotaping of classes, evaluation of course materials, and evaluation of instructor-graded student assignments if these methods were in place in formative evaluation involving colleagues as evaluators.

Research Question 2. What is the relationship between each of the methods of evaluation and each of its corresponding detractors?

Hypothesis 2: There will be a negative relationship between attitudes toward each method of evaluation and each of its corresponding detractors.

Research Question 3. What is the relationship between each of the methods of evaluation and each of the formative evaluation enhancers?

Hypothesis 3: There will be a positive relationship between attitudes toward each method of evaluation and each of the formative evaluation enhancers.

Research Question 4. What is the relationship between each method of evaluation and each of the individual/institutional improvement factors?

Hypothesis 4: There will be a positive relationship between attitudes toward each method of evaluation and each of the individual/institutional improvement factors.

Population

An examination of directories from catalogs of IAICU member institutions revealed that the 26 baccalaureate degree-granting institutions employed 2074 full-time faculty. That group constituted the population for this study.

Data Collection

A random sample of 750 (more than one-third of the population) of these faculty members was invited to participate in the study. Participants were requested to respond to a questionnaire about four methods of formative faculty evaluation and the other variables.

Instrumentation

The survey instrument, an original questionnaire, was based primarily on questions about methods of evaluation involving peers as evaluators that remained unanswered following a review of the literature. It was also based on questions that surfaced following a pilot study involving opinions of academic officers of IAICU member institutions (Keig, 1989). A study of the attitudes of faculty members from state-supported universities toward summative and formative evaluation by Britt (1982) was another important source of potential items to include on the questionnaire.

The main body of the questionnaire consisted of 29 content items in Yes-No-Not Sure and Likert-type formats. It also included eight demographic items in multiple-choice and completion-type formats.

Statistical Design

Descriptive statistics were computed for each of the Yes-No-Not Sure and Likert-type content items. These statistics were frequency counts and percentages for the Yes-No-Not Sure items. They were means and standard deviations, as well as frequency counts and percentages, for the Likert-type items. Frequency counts, percentages, means, standard deviations, and medians were calculated on the demographic variables of age and number of years of teaching experience at the college or university level. Frequency counts and percentages were computed on the dichotomous and categorical demographic variables (gender, academic rank, tenure status, and academic discipline).

The correlational procedures included the calculation of Pearson correlation coefficients among each of the four methods of evaluation and four or five variables in each of six categories. These six categories consisted of four classes of detractors (one for each method of evaluation), formative evaluation enhancers, and individual/institutional improvement factors.

The correlational procedures also included tests of statistical significance for Hypotheses 2, 3, and 4. These research hypotheses were accepted when the level of significance was less than .05 using non-directional tests.

Organization of the Study

The four chapters which follow were written to detail what has been presented in this chapter, for analysis of the data, and to offer conclusions reached on the basis of the findings. A review of related literature and research has been presented in Chapter II. Chapter III has detailed the procedures employed in the study. Statistical analyses of the data have been presented in Chapter IV. Chapter V has been devoted to a summary of the study, to recommendations for the practice of formative evaluation in higher education, and to recommendations for further study.

CHAPTER II

REVIEW OF RELATED LITERATURE AND RESEARCH

There has been general agreement that college teaching is not as good as it could, or should, be. There has also been widespread agreement that time-honored practices of faculty evaluation in higher education are not effective in improving instruction. In fact, Scriven (1980) described faculty evaluation as so "shoddy," "intellectually sloppy," and "slipshod," and such "a source of shame," that it is hardly surprising that teaching is rarely rewarded in an appropriate way (p. 7). Similarly critical of faculty evaluation, Soderberg (1986) described it as "simplistic" (p. 13), "primitive," and "without significant credibility" (p. 23). Likewise, Dressel (1976) described it as "generally quite limited, sporadic, and inadequate" (p. 333).

It has been said by those within the academy (and assumed by outsiders) that teaching is the central mission of colleges and universities (Arden, 1989; Blackburn & Clark, 1975; Britt, 1982; Dressel, 1976; Edwards, 1974). Essentially the same sentiment was expressed even more directly in Policy Perspectives, a publication of the Pew Higher Education Research Program (1989):

> Too seldom is collegiate teaching viewed for what it is: the business of the business--the activity that is central to all colleges and universities. . . . Teaching is the task that distinguishes colleges and universities, along with primary and secondary schools, from all other service agencies. (pp. 1-2)

At the same time that evaluation of instruction has been criticized as woefully inadequate, scholars have also charged that

far too much emphasis is being placed on the evaluation of scholarship and service (Miller, 1990; Soderberg, 1985). As a result, these scholars have concluded that many faculty members are disillusioned by (and are cynical about) the process and are not motivated to improve teaching (Soderberg, 1985, 1986; The Pew Higher Education Research Program, 1990).

Whether carried out formally or informally, effectively or ineffectively, summative evaluation has been (and continues to be) the principal focus of evaluation in higher education. For a long time, it was believed that summative evaluation would also function as an instrument in improving instruction. In recent years, however, the notion that there is a direct relationship between summative evaluation and instructional improvement has been challenged (Cross, 1986; Dressel, 1976; Soderberg, 1985). Since the late 1960s, programs of formative evaluation have been proposed (and occasionally implemented) as an alternative to summative evaluation in assisting faculty members who seek help in improving their teaching. Where they exist, these programs have been conducted mostly by the faculty or by instructional consultants (Cohen & McKeachie, 1981; Sorcinelli, 1984).

This chapter has been organized so that the ties between formative and summative evaluation and among peer, student, and administrative evaluation, as well as self-assessment, of teaching performance are highlighted. These relationships are highlighted, rather than minimized, so that readers are reminded that formative

evaluation and peer review are but components, though integral parts, of a comprehensive program of faculty evaluation. Consequently, the major divisions of this chapter have ranged from a general overview of comprehensive evaluation to the roles that might be played by peers, the methods that might be used in the formative evaluation of instruction, the factors which may take away from and enhance the process, and the benefits that may be derived from faculty participation in programs of formative evaluation of instruction.

Formative Evaluation of Instruction in the Context of Comprehensive Faculty Evaluation

The importance of making fair and impartial decisions about reappointment, promotion, tenure, and compensation cannot, of course, be disputed. Yet, many scholars have emphasized that an equally, if not more, compelling purpose for evaluating performance is to improve the quality of teaching. They have also suggested that efforts to improve instruction are important not only at the time personnel decisions are made (in the interest of students as well as faculty members), but also throughout the careers of faculty members in an effort to demonstrate the institution's commitment to teaching, to provide opportunities for faculty growth and development, and to provide some assurance to students that they are currently receiving (and will continue to receive) instruction of high quality (Menges, 1985).

Several scholars have at once criticized the lack of relationship between evaluation and teaching improvement efforts, and have advocated programs in which either the two functions are kept distinctly separate or inextricably linked. In that regard, Dressel (1976) wrote that "assessment activity must be broadly conceived as a basis for improvement, not for making personnel decisions. Evaluation can be linked to reward structures, but with recognition that improvement and development are the first concerns" (p. 374).

In a similar vein, Cross (1986) argued:

> The call for formative evaluation is loud and clear. Ironically, practically all the proposals and practices in assessment today involve summative evaluation. . . . There are few proposals for formative evaluation to show us how to improve education in process. . . . If we are to improve the quality of education, perhaps the most important question . . . to address is what decisions should be made to improve instruction . . . . How students are taught lies at the heart of quality education. It makes the difference between a lifelong learner and a grade grubber, between enthusiasm for learning and indifference to it, between an educated society and a credentialed one. (pp. 3-4)

Cross was certainly not satisfied that summative evaluation has resulted in instructional improvement, but was convinced that formative evaluation holds promise for improving the quality of teaching and student learning.

Formative Evaluation as a Component of Comprehensive Faculty Evaluation Programs

Nearly all writers in the field of faculty evaluation recommended the adoption of comprehensive programs of faculty evaluation (Aleamoni, 1981; Arden, 1989; Blackburn & Clark, 1975;

Brandenburg, Braskamp, & Ory, 1979; Dressel, 1976; Greenwood & Ramagli, 1980; Romberg, 1985; Sauter & Walker, 1976; Schneider, 1975; Scriven, 1980, 1983, 1985; Seldin, 1984; Smith, Hausken, Kovacevich, & McGuire, 1988; Soderberg, 1985, 1986; Spaights & Bridges, 1986; Stevens & Aleamoni, 1985; Swanson & Sisson, 1971). Comprehensive faculty evaluation has been described as a program in which several methods are employed and multiple constituencies are consulted in the process of gathering and interpreting data on the effectiveness of teachers. Programs of comprehensive evaluation have included both formative and summative components; have elicited information from students, faculty colleagues, academic administrators, and from faculty members being evaluated; and have included direct classroom observation, videotaping of classes, examination of course materials, evaluation of instructor-graded student assignments, and other methods in the process of gathering and interpreting data.

In insisting that it is critical to have a comprehensive program of faculty evaluation, as opposed to a more limited program, in place, Dressel (1976) observed that "no one method by itself is adequate. In fact, overemphasis on one method may do more harm than good. Various facets of the program can be examined by different and appropriate means of assessment" (p. 338).

Batista (1976) observed that "no technique or source of information is valid per se in evaluating college teaching.

Usefulness depends on the objectives to be reached" (p. 269). Likewise, Greenwood and Ramagli (1980) concluded that:

> None of the means of evaluating college teaching used alone seems to have a research base which indicates that it is a sufficiently valid measure of the teaching effectiveness of a given professor. Such a situation suggests the development of multiple data systems that are continuously validated and subject to ongoing empirical examination of the interrelationships existing between the different kinds of evaluation and instructional improvement data collected. (p. 681)

Yet, it has not always been possible, and probably never will be, for all colleges and universities to commit the necessary resources to implement comprehensive programs of faculty evaluation (Brandenburg, Braskamp, & Ory, 1979). In this event, scholars have recommended that the university community ask--and answer, after careful deliberation--three fundamental questions before conducting faculty evaluation. These questions were: (a) For what purpose is the evaluation being conducted?; (b) From what sources will the information be obtained?; Who will interpret the information gathered?; (c) What methods and procedures will be used to gather the information? (Bulcock, 1984; Cancelli, 1987; Craig, Redfield, & Galluzzo, 1986; Millman, 1981; Prater, 1983; Scriven, 1980; Soderberg, 1986). Further, these same scholars emphasized that a satisfactory answer to the first question must precede attempts to answer the other questions.

A number of reasons have been suggested for conducting faculty evaluations, but those mentioned most often have been for making personnel decisions and for improving instruction. Since the

purposes may be quite different, it has been suggested that formative and summative faculty evaluations may require different approaches to achieve the best results, as some methods and sources of information may be relevant for one purpose but irrelevant for the other. To illustrate this point, Centra (1975), Scriven (1980, 1983), and Sorcinelli (1984) argued that direct classroom observation, whether conducted by administrators or colleagues, is inappropriate for summative evaluation but is of great potential value in formative evaluation. Their criticism of using it for summative evaluation was based on its low inter-rater reliability, the time required to obtain a representative sample of teaching behaviors, and the likelihood that the teaching and learning environment is different when observers are present than when only the students and the professor are in the classroom. In one experimental study, Ward, Clark, and Harrison (1981) found that the presence of observers significantly altered approaches toward teaching, as professors were much more likely to involve students in the teaching and learning process when they knew that observers were in the classroom than when they were unaware of the presence of observers.

Another example of the potential conflict between summative and formative evaluation has also been cited. While academic administrators are usually considered indispensable in the process of summative evaluation, Edwards (1974) and McKeachie (1987) suggested that administrators may be a less credible source in

formative evaluation. These scholars reasoned that faculty members are not likely to ask for help from administrators in efforts to improve instruction, because these same administrators will eventually be called upon to make critical personnel decisions about them.

A Model for a Program of Comprehensive Evaluation, Including the Formative Evaluation of Instruction

Procedures for conducting summative and formative evaluation of programs and personnel at the dental school of the University of Maryland were described in detail by Romberg (1985). In this program, several methods of evaluation were employed and sources consulted, in five broad areas: (a) student evaluation of instruction, including both course quality and teacher effectiveness; (b) evaluation of the faculty for decisions regarding reappointment, promotion, and tenure; (c) the evaluation of each department, including the formative evaluation of its goals and objectives by the faculty as well as administrators, and the formative evaluation of the faculty by peers and department chairpersons; (d) faculty and student evaluation of administrators, including the dean, associate deans, and department chairpersons; and (e) evaluation of the goals and objectives of the school by students, faculty, and administrators.

The processes described in two of these five areas were quite conventional, in that students at most colleges and universities are afforded opportunities to evaluate courses and instructors,

and committees composed of faculty members and administrators at most colleges and universities are asked to make recommendations concerning personnel decisions. Systematic evaluation in the remaining three areas has been much less commonplace in higher education. The mechanisms which have been put into place to evaluate departmental goals and objectives and those of the school, as well as the formative evaluation of the faculty and administration by the faculty, have been employed infrequently.

A Rationale for Peer Review in Formative Evaluation

For this study, formative evaluation has been described as a component in comprehensive evaluation in which instructional improvement is the primary focus. It has been further stipulated that its objectives are to be accomplished by self-identification of strengths and weaknesses through interaction with, and with feedback from, one or more individuals who are knowledgeable about teaching and learning strategies and familiar with the teaching performance of the instructor being evaluated. Since scholars were convinced that there are advantages for keeping the summative and formative functions separate, they have recommended that the findings from formative evaluation should not be used in decisions regarding reappointment, promotion, tenure, or compensation.

McKeachie (1986) observed that scholars have not agreed, and probably never will, on a single theory by which effective teaching can be measured or evaluated. That conclusion notwithstanding, individual faculty members have probably developed, either explicitly

or implicitly, teaching styles consistent with their backgrounds, personalities, and experiences, and specific theories of teaching and learning (Bulcock, 1984; McKeachie, 1986). For that reason, scholars have suggested that informed peers--who know their colleagues personally and professionally, who are familiar with effective (and ineffective) practices in the styles adopted by their colleagues, and who recognize that effective teaching is contingent on a number of complex factors, only some of which are controlled by instructors--may be forceful catalysts in the process of instructional improvement.

A number of scholars criticized the inflexibility built into most programs of faculty evaluation (Bulcock, 1984; Cancelli, 1987; Scriven, 1980, 1983). To remedy this situation, Cancelli (1987) envisioned a system of peer evaluation that:

> makes minimal assumptions regarding how instruction should occur. It is left to the professional judgments of the professors to determine how they wish to develop and teach their courses. The system only requires that they be within the bounds of acceptable practice, broadly defined, that they do what they say they do, and that there be a cogent rationale for their choices. Thus, the review of each professor is unique and requires decisions based on disparate and often idiosyncratic bits of information. The use of judgments by peers provides a method that is flexible enough to adjust to the unique data base generated in each review. (p. 12)

While formative evaluation of instruction cannot address all of the problems inherent in the system of faculty evaluation, many scholars and practitioners have agreed that it is one promising means for improving the quality of teaching. However, these same scholars and practitioners have also warned that peer review must

be carefully planned and implemented if it is to be an effective component in the process of formative evaluation.

Models of Formative Evaluation Involving Colleagues as Evaluators

A number of programs of formative evaluation in which peer review is an integral component have been developed by scholars and researchers. While many of these programs have proceeded no further than the theoretical stage, some of them have been implemented by elementary and secondary schools and by colleges and universities in the United States and abroad.

Roper, Deal, and Dornbusch Model. A carefully conceived formative evaluation model was developed and put into place by Roper, Deal, and Dornbusch (1976). This plan included seven interdependent stages: (a) identifying the participants and determining which of them will work together; (b) defining and clarifying the teaching and learning objectives of each participant; (c) the setting of the criteria by which the performance of each participant will be evaluated, based on the outlined objectives; (d) assessing the quality of the performance of each participant through the use of direct classroom observation, videotaping of classes, evaluation of course materials, evaluation of instructor-graded student assignments, and/or other methods; (e) appraising the strengths and weaknesses of the performance; (f) communicating the results of the evaluation, normally through direct interaction between/among evaluator(s) and the faculty member whose

performance is being evaluated; and (g) developing a plan for improvement based on self-assessment as well as on the results of the process.

Roper, Deal, and Dornbusch (1976) recommended that the membership of the working groups be determined by the program participants, believing that this method of selection is preferable to random selection or arbitrary assignment. They concluded that mutual respect and trust between/among participants, and compatible educational philosophies that can be controlled for through self-selection, outweigh advantages offered by other methods of selection. Agreeing that participants should determine membership of the working groups, Heller (1989) warned that an attempt to impose membership might result in groups that "go through the motions," but make little or no legitimate attempt to improve the quality of teaching.

In the Roper, Deal, and Dornbusch model, and in other formative evaluation models, participants have been encouraged to identify the worth of their own objectives (either individually or collectively), to determine the methods by which the data will be gathered, and to establish the criteria by which strengths and weaknesses will be evaluated. Allowing participants to make these fundamental decisions was seen as a means for developing faculty ownership in the program (Heller, 1989). While performance is evaluated largely or exclusively by direct classroom observation in many models, a number of methods for gathering information were

recognized as viable by Roper, Deal, and Dornbusch (1976). These researchers recognized that a variety of effective teaching styles and methods for evaluating them exist.

The method for communicating the feedback from formative evaluation has become largely standardized. After the agreed upon methods of evaluation have been completed, the group members have met together to discuss the findings. At these sessions, each group member, including the faculty member whose performance is being evaluated, has been encouraged to provide an assessment of the items under consideration. Faculty members in the Roper, Deal, and Dornbusch study reported that criticism was most often presented as suggestions for alternative techniques, rather than as mandates or absolutes. The researchers reported that comments "encompassed virtually every aspect of classroom activity. Teachers learned not only about their own performance but about the overall climate of their classrooms" (Roper, Deal, & Dornbusch, 1976, p. 62).

The planning of the programs for improvement has evolved directly from the feedback sessions, from self-assessment of other aspects of the process, and from self-reflection on performance in general. At these planning sessions, participants have worked collectively to determine the strategies to be employed in efforts to improve performance, the resources required to accomplish these objectives, and the means by which these instructional improvement efforts will be evaluated.

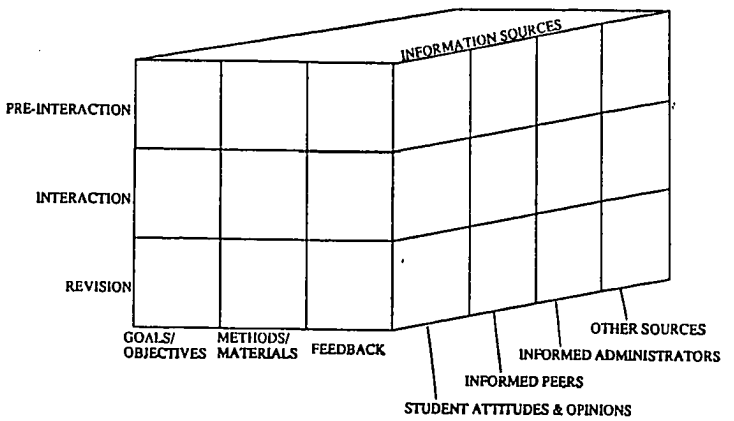
While participants in the Roper, Deal, and Dornbusch study were enthusiastic about the program, the model apparently has not been adopted by other elementary or secondary schools or by colleges or universities. The reasons for this have not been addressed by researchers.

Soderberg Model. A second model of comprehensive faculty evaluation, adapted from a number of teaching and learning theories, was developed by Soderberg (1986). The three dimensions of the model have been diagramed to show how elements from each dimension relate to each other (Figure 1).

Soderberg described the first dimension as consisting of a series of interdependent conceptual events that instructors consider in setting goals and objectives, in determining how to accomplish these goals and objectives, and in selecting methods for evaluating their success in meeting the goals and objectives. The second dimension was described as another interdependent process of more concrete instructional events: pre-interactive events (planning for what should occur during the process of instruction), interactive events (the actual delivery phase), and a post-interactive phase (the process of revision, in which reflection and further planning guide efforts to improve the delivery of instruction).

As the third dimension, Soderberg identified several constituencies which may be able to provide information about the strengths and weaknesses in the instructional process. While

Figure 1. Soderberg's three-dimensional model of faculty evaluation.



Note. From "A Credible Model: Evaluating Classroom Teaching in Higher Education" by L. O. Soderberg, 1985. Instructional Evaluation, 8(2), p. 20. Reprinted by permission.

students and informed peers were listed as primary sources, informed administrators and alumni were identified as important secondary sources of information.

The Role of Colleagues in the Formative Evaluation of Instruction

In formulating a model for comprehensive faculty evaluation, Soderberg (1986) recognized that informed peers are potentially a primary source of information in the evaluation of the teaching performance of their colleagues. Many other scholars also suggested that faculty members have expertise that should be utilized in formative, as well as summative, evaluation of instruction (Aleamoni,

1981; Batista, 1976; Centra, 1975; Cohen & McKeachie, 1981; Fitzgerald & Grafton, 1981; Gunn, 1982; Wilson, Dienst, & Watson, 1973). However, Cohen and McKeachie (1981) emphasized that peer review of isolated elements of teaching performance "are not necessarily valid indicators of effective teaching when used by themselves, but they may be helpful when used in conjunction with other evidence" (p. 147).

Elements of Teaching That May Be Better Evaluated by Colleagues Than by Students or Academic Administrators

Batista (1976) and Cohen and McKeachie (1981) provided the most exhaustive analyses of the roles that might be played by peers in the evaluation of instruction. Cohen and McKeachie classified these roles into four broad, if somewhat overlapping, categories: (a) elements of course design, including goals and course content; (b) instructional methods and materials; (c) evaluation and grading practices; and (d) the integration of information from all sources.

In the category of course design, these researchers cited the following elements: (a) mastery of course content; (b) appropriateness of course objectives; (c) selection of course content (knowledge of what must be taught); (d) organization of the course; (e) coverage of appropriate content; (f) incorporation of recent scholarship into selection of content; and (g) suitability of student assignments in meeting course objectives. Soderberg (1986) noted that informed peers may also the best judges of the suitability of objectives for particular groups of students and the appropriateness of the rigor of the course in its contextual environment.

In the category of instructional methods and materials, Cohen and McKeachie (1981) cited five areas in which colleagues might capably judge the effectiveness of faculty members: (a) suitability of methods of instruction in meeting course goals; (b) appropriateness of the reading list for the course; (c) reasonableness of the amount of time required of students for completing readings, written assignments, and other projects; (d) appropriateness of handouts and other instructional materials in facilitating learning; and (e) the suitability of various types of media in meeting course objectives. Batista (1976) included a related item: the application of appropriate methodologies for teaching specific content.

In assessing the devices employed by professors for evaluating student assignments and their grading practices, Cohen and McKeachie recommended that colleagues judge: (a) the length and difficulty of examinations; (b) the coverage given to higher-order, as well as lower-order, cognitive processes on examinations and other assignments; (c) the time and effort required of students to complete written assignments and other projects; and (d) the specificity on how grading practices are explained to students. Soderberg (1986) recommended additional areas in which peers might become involved in evaluating faculty assessment of student performance: the relationship of evaluation instruments to course objectives and procedures, the usefulness of the evaluation to students in the learning process, and the relationship between the awarding of

course grades to the grading system communicated to students (presumably) early in the term.

Scriven (1985) commented on two related items, test construction and the actual grading of examinations. In suggesting that colleagues with specific expertise in test construction and evaluation become involved in assessing the competencies of colleagues in these areas, Scriven observed that "few teachers . . . have ever had their tests and scoring keys looked at against minimum standards of professional competence, if indeed they have ever heard of such standards; and those that have been looked at present a very depressing picture" (p. 32). Scriven also discussed a number of specific means by which test construction and grading might be improved with the assistance of psychometricians.

Cohen and McKeachie (1981) indicated that colleagues are ideally equipped to integrate and interpret information gathered from various methods of evaluation and from other sources. In the evaluation of instruction, these researchers suggested that informed peers consider the following factors: (a) the interpretation of student ratings in light of circumstances under which the course was taught (e.g., large versus small enrollment in classes, required versus elective courses, and a number of other contingencies which may affect student ratings); (b) the criteria used in evaluating instruction; and (c) the weighting of the criteria used in determining the effectiveness of teaching in specific contexts. Batista (1976) listed a number of other factors that might be

considered by colleagues in the integrative process of instructional evaluation: (a) faculty members' own evaluations of their teaching; (b) faculty members' own evaluation of knowledge of specific content areas in the field as a whole; (c) informal course evaluations conducted by instructors with their students; (d) alumni ratings of faculty members; (e) student achievement in courses; and (f) interviews with groups of students or individual students.

Craig, Redfield, and Galluzzo (1986) and McKeachie (1986) recommended that the integrative process include a study of the relationship between students' explanations of their responsibility in teaching and learning, and their evaluations of courses and instructors. The scholars were convinced that information provided by these comparisons would be a valuable tool in assessing the quality of teaching and student learning.

Summary

When faculty members become involved as evaluators in programs of instructional improvement, writers in the field have suggested that it might become a mechanism by which teachers work together to improve the quality of instruction. Unlike summative evaluation, where all faculty members are evaluated according to common criteria, Menges (1985) observed that formative evaluation:

> proceeds differently from person to person and from setting to setting. Its initiation is usually from within, although certain external circumstances support it more effectively than others. . . . These conditions include exposure to relevant theory, provision for practicing the new approaches and receiving feedback in the practice, and opportunities to be coached while applying the new approaches in the field. (p. 9)

Scholars have also agreed that faculty members have areas of expertise about teaching and learning that should be employed in faculty evaluation. In the formative evaluation of instruction, scholars have suggested that this involvement might come through direct classroom observation, the evaluation of videotapes of classes, the evaluation of course materials, and the evaluation of instructor-graded assignments. These methods of evaluation have been chosen as the focus for the next section of this chapter.

Methods of Evaluation

A number of methods have been used to evaluate the faculty in higher education. As indicated earlier, scholars have come to believe that some methods are more appropriate for summative evaluation than for formative evaluation, and vice versa. The methods that have been mentioned frequently as most appropriate for formative evaluation are direct classroom observation, videotaping of classes, the evaluation of course materials, and the evaluation of instructor-graded student assignments.

Direct Classroom Observation

While many scholars have expressed serious reservations about the use of direct classroom observation in summative evaluation, nearly all of them have recommended that it be employed as one means of evaluation for course improvement purposes. They have been especially supportive of this method when it is to be carried out by fellow faculty members.

Support. Scholars who support direct classroom evaluation conducted by the faculty have argued that it should be employed because faculty members have expertise in the process of teaching and learning that is not possessed to the same degree by either students or administrators (Cohen & McKeachie, 1981; Soderberg, 1986; Sorcinelli, 1984). Some of them also suggested that it is a vital component of the process, because it is the only way in which some aspects of teaching can be judged adequately.

Seldin (1984) concluded that there are specific reasons for involving peers in the evaluation of fellow faculty members. When employed for formative evaluation, Seldin believed that classroom visitation poses minimal threat to those who are being evaluated, can be handled informally between or among two or more participants (as well as more formally, if so desired), can be made adaptable to different modes of instruction, and is a means of developing deeper trust between or among the participants.

In regard to improving the quality of instruction through peer evaluation, Hart (1987) observed:

> To improve, teachers need the help and support of other teachers. Teachers need to consult regularly, over an extended period, with other teachers. Teachers need to observe other teachers at work, be observed by them in return, and share their observations, reflections, and recommendations. (p. 15)

But the same thinker cautioned that direct classroom observation may not be "easy, comfortable, simple, or quick in results" (p. 15).

A number of scholars, including Braskamp (1978) and Heller (1989), recommended that certain preconditions be met before direct classroom observation involving colleagues as evaluators is put into place. They suggested that: (a) participation be voluntary and, if possible, initiated by the faculty; (b) those who will evaluate and in turn be evaluated be trusted and respected by each other; (c) classroom visits be by invitation only (there should be no surprise visits); (d) participants determine in advance what aspects of their teaching will be evaluated; and (e) participants determine in advance how the strengths and weaknesses of their performance will be documented.

Writers in the field disagreed on whether participants should come from the same (or related) academic disciplines or from different disciplines. Scriven (1985) argued that they should come from the same discipline, because knowledge and transmission of specific course content can be judged effectively only by those who are steeped in that discipline. Expressing the opposite view, Heller (1989) and Shatzky and Silberman (1986) suggested that participants might come from different disciplines, as colleagues who are largely unfamiliar with course content may be more likely to understand difficulties experienced by novices in the discipline.

The models on which most programs of direct classroom observation have been based are the clinical supervision models of Goldhammer (1969) and Bergquist and Phillips (1975). The models described by these scholars have included the following three stages:

(a) a pre-observation conference in which the instructor's goals and objectives, teaching strategies, classroom procedures, and the methods for evaluating them are agreed on; (b) the observation of one or more class sessions; and (c) a post-observation conference in which the instructor and the observer meet together to discuss what occurred during the classroom visitation. Goldhammer (1969) and Bergquist and Phillips (1975) emphasized the importance of providing opportunities for feedback about specific teaching and learning incidents following the visit. Skoog (1980) and Sorcinelli (1984) provided even more detailed descriptions of what might occur at each of the stages of the process.

While several programs in which peers have observed a colleague's classroom have been presented in the literature (some of which will be described later in this section), much less has been written about the specific events that might be observed. A particularly cogent discussion of six interrelated, if not exhaustive, categories of these events was provided by Hart (1987).

1. The physical-temporal setting. Hart noted that the:

> time of day, room size and shape, air (or the lack of it), light (or dark), surrounding noise, furnishings, apparatus, and clutter [affect] how the people of the event use or misuse this environment: their uses of space, access, positioning, distance, mobilities. . . . [While] the teacher may need few reminders of the ecology, the reactions of the observer may well help to understand and use it better. (pp. 17-18)

2. The logics of the situation. Hart indicated that each class has:

> its intellectual structures, orders, sequences, its texture of governing ideas, its proportions, connections, transitions, planned or not. Since teachers regularly signal to the class what these are, others (ill-advisedly, I think) take them for granted. Teaching is, among other things, a composing process. . . . The observer can, at least, keep track of the structures or logics that are communicated, and report them back--occasionally to the teacher's surprise. (p. 18)

3. The linguistic dimension. Hart noted that certain types and levels of language are employed by teachers and students, and concluded that:

> they are sometimes similar, sometimes quite distinct--even separate or divisive. . . But not many teachers in my experience are aware of the languages they use. The observer can hear and report the relative degrees of difficulty, formality, technicality, the dominant syntactic forms. . . Every class session has its rhetoric: certain forms and methods that are used to achieve certain ends--informative, explanatory, persuasive. . . To carry out these complex aims, the teacher uses certain tactics and methods: assignments, exercises, demonstrations, examples, analogies, and motivational appeals. The observer can learn to observe and report the ends and the appropriateness of the means. (p. 18)

4. The dramaturgical setting and sociological context. Hart observed:

> We are all sufficiently familiar with dramaturgy to be useful observers of how members of a class play their roles and how they interact. We can record such phenomena as pacing, voicing, nonverbal behavior and communication, the class dynamic, its degree of intensity and involvement. . . We can observe how the teacher uses authority (or power) and what kinds, interpret the politics of the class, the directions and commands, invitations, judgments, rewards, and threats. (p. 19)

5. The curricular context. While the curricular context probably cannot be directly observed, an observer can make certain inferences about the class's relationship to:

> larger designs, other courses and areas of study, other disciplines, levels of learning and development, academic goals and values, extramural preoccupations and influences. No class is an island. What uses are made of such foreign relations, and how many, can be observed. What uses and how many should be made is a legitimate issue of strategy and priority. (p. 19)

6. The effects of teaching. Hart stressed that the outcomes of teaching are what really matter, but noted:

> Most teachers unwittingly cling to the assumption that time needed for teaching leaves no time in class for finding out what is being learned. . . The observer can only try to catch the clues and report them, and try to help the teacher find and use more adequate ways of discovering what has been learned. (p. 19)

Hart provided guidelines that might be useful to faculty members (and administrators) who evaluate the teaching performance of professors through direct classroom observation. While this listing seems imposing, even overwhelming, it was also suggested that it be used as a conceptual tool by individual faculty members during the pre-observation phases as a guide to identifying specific areas of teaching which are of concern to them.

Criticism. Despite a call for peer involvement in direct classroom observation of the teaching faculty, detractors have contended that there are limits to what can be observed and that there are several potential threats to its reliability and validity. Wood (1977, 1978) found that the process can be biased in the

following ways: (a) association (faculty who have close professional and/or personal associations are more likely to rate each other higher than those with whom they are associated less frequently); (b) visibility (faculty whose offices are located near the central office are more likely to be rated higher than those whose offices are located in more remote areas); (c) lack of independence between ratings for teaching and research (faculty who are rated high on research are also likely to be rated high on teaching); (d) lack of independence between ratings for teaching and service (teachers who are rated high on service are also likely to be rated high on teaching); (e) lack of independence between ratings for teaching and the number of credit hours taught (faculty who teach heavier class loads are more likely to be rated higher than those who teach lighter loads); (f) lack of independence between ratings for teaching and number of graduate courses taught (faculty who teach larger numbers of graduate courses are more likely to be rated higher than those who teach fewer graduate courses); (g) faculty who teach elective courses are more likely to be rated higher than those who teach required courses; and (h) lack of independence between academic rank and ratings for teaching (faculty who are at the higher professional ranks are likely to be rated higher on their teaching than those at lower ranks).

Centra (1975) identified other potential threats to the reliability and validity of peer evaluation. Centra noted that there is low inter-rater reliability, that faculty are more generous

in their ratings than students, and that attaining a large enough sample of classroom behaviors in order to make accurate generalizations may be prohibitively time-consuming. While concluding that these problems are difficult but not insurmountable, this researcher cautioned that these factors should be carefully considered before direct classroom observation using peers as evaluators is put into place. Centra's concerns have also been raised by other researchers, and have been important considerations in methodological aspects of this dissertation study.

Most of the concerns of Bergman (1979, 1980), Centra (1975), and Wood (1977, 1978) regarding the reliability and validity of direct classroom observation were made in connection with summative evaluation. Nevertheless, some of the same concerns have sometimes been raised in regard to formative evaluation. However, Centra (1975, 1986) observed that these factors may not be of so much concern in formative evaluation, as the participants are largely volunteers who are actively seeking feedback from trusted and respected colleagues, none of whom seek to gain from providing inaccurate information. In short, it has been said that faculty members who sincerely want to improve their teaching have little to gain from biasing the results.

Programs in place. Programs of formative evaluation in which direct classroom observation is carried out by the faculty have been implemented at Evergreen State College (Elbow, 1980), San Jose State University (Galm, 1985), Texas Tech University (Skoog,

1980), the University of Birmingham, England (Mathias & Rutherford, 1982a, 1982b), the University of Cincinnati (Sweeney, 1976; Sweeney & Grasha, 1979), the University of Kentucky (Cowen, Davis, & Bird, 1976), the University of Maryland (Romberg, 1985), the University of New York--Cortland (Shatzky & Silberman, 1986), and the University of South Carolina (Bell, Dobson, & Gram, 1977). While there were a number of similarities among these programs, there were also significant differences.

The common threads among them included: (a) voluntary participation, (b) faculty control over the teaching and learning events to be observed and commented on, (c) development of open lines of communication among the participants in the presentation of feedback, (d) a strategy in which praise and criticism are carefully balanced, and (e) a policy in which results are kept strictly confidential. Otherwise, each program was different in some respect from the others.

Most of the programs incorporated elements from the clinical supervision models of Goldhammer (1969) or Bergquist and Phillips (1975). A typical program of this type was put into place at the University of Cincinnati (Sweeney, 1976; Sweeney & Grasha, 1979). It involved collaborative teams of three members who worked together over a period of one semester. In the first phase of the program, each participant independently listed a number of instructional goals and the teaching strategies to be employed for the class period to be observed. In the second phase, team members met

together to discuss and clarify the lists of each participant, to determine the focus of each of the observations, to specify the techniques to be employed in gathering and reporting the data, to outline the methods by which the feedback was to be communicated, and to schedule the first round of classroom visitations.

The actual classroom observations were conducted during the third stage of the program. During these classroom visitations, the observers were asked to confine their formal critiques to the areas agreed on during the pre-observation stages. Following each classroom visit, a post-observation conference was scheduled. At each of these conferences, the observers recalled specific incidents, pointed out positive aspects of the teaching performance, and offered suggestions and alternative approaches in instances where they believed that performance could be improved. At this session, the faculty member who had been observed was encouraged to respond to the feedback by asking questions, seeking clarification, and commenting on points made by the observers. At this same session, the faculty member was asked to plan a program of improvement based on the observations and from self-assessment of the process. Finally, a date for a second classroom visitation was scheduled.

After the group had completed a full round of observations, the three members of the group met again to assess the process, including their perceptions of successes, failures, and/or interpersonal problems hampering them. Following this final phase of the program, the process was repeated, beginning with the

goal-setting phase and continuing through the process evaluation phase.

A somewhat different approach was taken at Evergreen State College (Elbow, 1980). At Evergreen, selected faculty members conducted intensive observations, over a period of several days, and conducted detailed interviews with program participants and their students. At the University of Kentucky (Cowen, Davis, & Bird, 1976), faculty members offering courses and seminars for the first time were asked to present lectures or seminar topics to their colleagues for feedback. In yet another example, two faculty members at the University of New York--Cortland (one from English and the other from chemistry) attended each other's classes for an entire semester. Both of them completed all assignments and tests, took part in student study groups, and conducted informal discussions with students about the teaching and learning that was taking place in the courses they were taking (Shatzky & Silberman, 1986).

Evaluation of programs. In all instances, researchers reported that faculty members believed that their teaching and student learning had improved as a result of the feedback provided by colleagues. However, evaluation of these programs has rarely been systematic. In two experimental studies, modest, though statistically significant, improvement in student ratings of faculty members and in student learning were reported (Erickson & Erickson, 1979; Hoyt & Howard, 1978).

Direct classroom observation has been employed in several different ways at colleges and universities in the United States and abroad in the 1970s and 1980s, but it has not been the only method of formative evaluation employed. Other methods have included videotaping of classes, evaluation of course materials, and evaluation of instructor-graded student assignments.

Videotaping

The use of videotaping of classes in the formative evaluation of faculty in higher education has received much less attention in the literature than direct classroom observation. While videotaping has occasionally been used as a component in faculty evaluation, no systematic studies of videotaping involving colleagues as evaluators have been reported (Centra, 1977). Still, a number of scholars have recommended ways in which this method might be employed in improving the quality of teaching. They have also offered a number of caveats concerning its use.

Dressel (1976) observed that videotaping might be used to illustrate "weaknesses in delivery, in expression, in emphasis, and in attention to students--all of which can be improved" (p. 351). Smith, Hausken, Kovacevich, and McGuire (1988) added that, in addition to weaknesses in presentation, videotaping can provide useful clues to student responses to what was presented and to how the instruction was provided.

Craig, Redfield, and Galluzzo (1986) envisioned an even greater use for videotaping. They recommended that it be used in "stimulated

recall interviews," a self-report process in which a videotape of a class is played and stopped periodically for students to report what they thought about and how they reacted to specific incidents at strategic points during the class session. The same thinkers suggested that peers can best assist a colleague in interpreting the information provided by the stimulated recall interview.

Nearly all writers on videotaping commented on the dangers of its self-confrontational nature (Brandenburg, Braskamp, & Ory, 1979; Braskamp, 1978; Brock, 1981; Craig, Redfield, & Galluzzo, 1986; Dressel, 1976; Seldin, 1984; Smith, Hausken, Kovacevich, & McGuire, 1988). In that regard, Seldin (1984) noted that viewing a videotape can be a "painful and an illuminating experience" (p. 351).

Brandenburg, Braskamp, and Ory (1979) expressed well the sentiments of many writers in this area by concluding that findings from videotapes are "especially personal and descriptive; viewing a videotape with a colleague is preferable to only the instructor viewing it because the colleague can share his/her insights, can provide support in this confrontational experience, and suggest improvements and changes" (p. 12). The conclusion of these writers, then, was that viewing of videotapes of classes can be a valuable experience if handled judiciously, but can be damaging if not handled with care.

Evaluation of Course Materials

While scholars of faculty evaluation in higher education have generally agreed that informed peers are ideally suited to evaluate

their colleagues' course materials, Cohen and McKeachie (1981) and Seldin (1984) observed that few colleges and universities have integrated this procedure into the evaluation process. A number of reasons have been offered for employing this form of evaluation so infrequently, but no firm conclusions have been reached.

Besides Seldin's (1984) simple explanation that its limited use may be merely one of oversight, Centra (1986) offered three other, more fundamental, reasons: (a) course materials are so personal and subjective that faculty members are not willing to open course materials up to the same close scrutiny that they give colleagues in the critiquing of manuscripts for publication; (b) the time required to review their colleagues' course materials can be better spent on research, where the extrinsic rewards are usually greater; and (c) it is not worth the time because course materials are read "only" by students, while published research is there for all to read and evaluate.

Despite the limited use of evaluation of course materials in the formative evaluation of instruction, several writers catalogued a number of elements of course design and materials that could be evaluated by informed peers (Aleamoni, 1981, 1984; Brandenburg, Braskamp, & Ory, 1979; Braskamp, 1978; Cancelli, 1987; Centra, 1986; Cohen & McKeachie, 1981; Dienst, 1981; Eckert, 1950; McCarthey & Peterson, 1988; Scriven, 1980, 1983, 1985; Smith, 1985; Smith, Hausken, Kovacevich, & McGuire, 1988). Cohen and McKeachie

(1981) presented many of the aspects of course design and teaching materials that might be evaluated by colleagues (Figure 2).

When course materials are to be evaluated by colleagues for instructional improvement purposes, Smith, Hausken, Kovacevich, and McGuire (1988) suggested that the procedure be similar to that employed when direct classroom observation is used for formative evaluation. Specifically, they recommended that the process include: (a) independent evaluations of the syllabus, textbooks, supplementary readings, and handouts by a panel of reviewers, based on predetermined criteria (e.g., currency, relevancy, and/or accreditation standards); (b) a meeting in which the reviewers discuss their independent findings and attempt to arrive at consensus; (c) a post-review conference in which the course instructor and the reviewers meet to discuss the findings, to clarify misconceptions, and to provide feedback; (d) the writing of a summary of the findings and recommendations; and (e) an extended period of time in which the instructor can consider the recommendations, and then either plan for, implement, and evaluate changes or provide a rationale for a decision not to implement the recommendations of the reviewers.

The evaluation of faculty members' course materials in the process of formative evaluation, whether done by peers or administrators, has not been approached systematically at most colleges and universities, if it has ever been used. Another method

Figure 2. Elements of course design and course materials for evaluation by peers

Course goals/course content

- Goals are appropriate for the specific course
- Basic subject matter content is covered in the course
- Current course material is presented
- Work requirements for the course are appropriate

Instructional methods and materials

- Method of instruction is suitable to meet course goals
- The reading list is appropriate for the course
- Reading assignments require a reasonable amount of time and effort
- Handouts and learning aids are used appropriately
- Media materials (e.g., films, videotapes, audiotapes) are used in an appropriate manner

Evaluation and grading devices

- Exams cover the higher level goals of the course
- Exams are reasonable in length and difficulty
- Written assignments make students think
- Written assignments require reasonable time and effort
- The criteria for grading are appropriate

Note. From "The Role of Colleagues in the Evaluation of College Teaching" by P. A. Cohen and W. J. McKeachie, 1981. Improving College and University Teaching, 28, p. 151. Copyright 1981 by Heldref Publications. Reprinted by permission.

of assessment, evaluation of instructor-graded student assignments, has been employed even less frequently.

Instructor-Graded Student Assignments

Despite the fact that the evaluation of instructor-graded student assignments (tests, term papers, individual or group projects, homework, etc.) has been employed infrequently in higher education, a number of scholars commented on the benefits that might be derived from, and the problems associated with, using this method as one component of a comprehensive program of evaluation (Bryant, 1967; Cohen & McKeachie, 1981; McKeachie, 1986; Scriven, 1980, 1985). In general, these scholars concluded that this method is more useful for course improvement purposes than it is for decisions regarding reappointments, promotion, tenure, and compensation.

Thoughtful comments concerning an instructor's responsibilities in facilitating student learning, which might be viewed as a rationale for including peer review of instructor-graded student assignments as part of the process of evaluation, have been articulated by Dressel (1976).

> [It is an instructor's responsibility] to provide the student with satisfaction through a sense of progress. The responsibility requires pointing out to the student both successes and deficiencies . . . Evaluation for feedback and motivation is an essential component of good teaching. [Little can] replace the personal commendation of an admired teacher. Praise or the regard of others is a potent motivator.
>
> [Yet] an indispensable aspect of learning is the recognition and admission of error, combined with the ability to profit from error. Failure must

> come to be regarded as a challenge, not as a disabling or uncorrectable event which impedes further progress. (p. 343)

In this statement, Dressel implied that there is a delicate balance between praise and constructive criticism, and that teachers often do not achieve this balance. McKeachie (1986) observed that instructors rarely exploit the student examination as a vehicle for providing meaningful feedback to students, and suggested that the situation might be improved by asking colleagues to review instructor-graded tests and other assignments as a means of determining the adequacy of the feedback to students.

Scholars have also mentioned other potential benefits that might be derived from peer review of instructor-graded student assignments. For example, Cohen and McKeachie (1981) observed that this method of evaluation might provide a way of determining students' higher-order cognitive gains (i.e., it might show how students apply principles learned in a course in solving problems, or their breadth of understanding in relating what they have learned to other courses and disciplines).

Scholars have warned that assessing student achievement on the basis of instructor-graded student assignments (or even standardized tests), be it by colleagues or administrators, involves difficult psychometric problems. In this regard, Centra (1986) concluded that "there is at this time no evidence that these assessments will be valid or reliable" (p. 4). The same author also noted that this method of evaluation is especially

problematical, because there are many other variables, besides teaching, that affect student learning. Nevertheless, Cohen and McKeachie (1981) concluded that "colleagues, who have a sense of typical student performance, are in the best position to judge the instructional impact on students. As of now, though, such judgments are qualitative in nature and can be best used for supplementing other data" (p. 151).

There has been little systematic study of instructor-graded student assignments in relation to instructional evaluation. Why this has not occurred, when those who have thought about this method research support its utility, was considered when the research questions, hypotheses, and questionnaire for the present study were formulated.

Detractors

It was noted earlier that direct classroom observation, videotaping of classes, evaluation of course materials, and evaluation of instructor-graded student assignments are not common components of instructional improvement programs in higher education. A number of reasons have been cited for the unwillingness of faculty members to implement and participate in formative evaluation when these methods are employed. Some of the reasons were based on practical considerations; others were more philosophically based. The most substantive of the issues, those considered in this study, were: academic freedom, the nonrepresentativeness of the teaching

and learning situation when evaluated by colleagues, the accuracy of methods of evaluation, and the subjectivity of the measure.

Academic Freedom

Several scholars considered the issue of academic freedom vis-a-vis evaluation conducted by direct classroom observation (Cross, 1986; Eckert, 1950; Edwards, 1974; Farmer, 1976; Hart, 1987; Roper, Deal, & Dornbusch, 1976; Sweeney & Grasha, 1979). While acknowledging that there is often the perception that direct classroom observation by peers (or administrators) violates a professor's academic freedom, those who have thought about this issue have generally concluded that academic freedom is not compromised by direct classroom observation, because the right of faculty members to determine and control what is taught is not circumscribed by the process of observation. Nevertheless, these thinkers have agreed that gaining the support of the faculty for this and other methods of evaluation involves overcoming faculty concerns over the issue of academic freedom.

As early as 1950, Eckert advocated the use of direct classroom observation involving colleagues as evaluators, but warned that it should not be put into place "without making exceedingly careful advance preparations" (Eckert, 1950, p. 67). This advanced planning included eliciting faculty support so that direct classroom observation would not be seen as an infringement on their academic freedom.

Nonrepresentativeness of Teaching and Learning Situations When Evaluations Are Conducted

A second factor which may have contributed to the unwillingness of faculty members to develop, implement, and participate in formative evaluation conducted by colleagues was their belief that the methods used are not valid measures of typical teaching and learning situations. The most common criticism of this type has come in connection with direct classroom observation when it is used as the only method in evaluating teaching performance. When it is employed, the usual practice has been to have two or three observers visit a class on one or two occasions. Most scholars concluded that such a procedure is not reliable enough to generalize about a faculty member's teaching performance (Centra, 1975; Prater, 1983; Scriven, 1980; Soderberg, 1986). In strongly objecting to the use of direct classroom observation as the primary means for making personnel decisions, Stodolsky (1984) wrote that classroom visits:

> are unlikely to be fair because any given observation will not be representative of the range of teaching behaviors used by a [teacher]. Evaluators are mistaken if they assume they are observing typical behaviors of [teachers] with the usual procedure. (p. 17)

Despite that admonition, the same thinker also noted:

> One might use observations as one type of information in conjunction with other materials that could provide a more rounded assessment of a teacher.
>
> In formative evaluation, direct observation may be very appropriate if too much is not made of any given observation. Direct observations can provide

> useful occasions for dialogues with supervisors and colleagues. Specific occasions are what teaching is all about, and may provide a very appropriate focus for discussing improvement. Discussions and suggestions that follow observation of a teacher may be even more helpful if it is recognized that he or she might teach differently in different situations. Rather than assuming that one knows a teacher well after a limited set of observations, one might rather acknowledge the incompleteness of that knowledge. (p. 17)

While recommending direct classroom observation as a means for improving classroom teaching, Hart (1987) reminded readers:

> No outsider, no occasional visitor to the ongoing intellectual community of class, can hope to understand very fully the internal processes, the codes and interactions, of that community. No mere observer can fully understand the roles of participation in the class. And this is an important limitation, for the roles of participation control what happens in a class and how such happenings are perceived and responded to. (p. 16)

Most of the same criticisms and observations made in regard to direct classroom observation can also be made about videotaping of classes, evaluation of course materials, and evaluation of instructor-graded student assignments. Scholars suggested that the most complete assessment of teaching performance is made when several methods of evaluation are used in conjunction with each other (Aleamoni, 1981; Arden, 1989; Blackburn & Clark, 1975; Bradenburg, Braskamp, & Ory, 1979; Dressel, 1976; Greenwood & Ramagli, 1980; Romberg, 1985; Sauter & Walker, 1976; Schneider, 1975; Scriven, 1980, 1983, 1985; Seldin, 1984; Smith, Hausken, Kovacevich, & McGuire, 1988; Soderberg, 1986; Spaights & Bridges, 1986; Stevens & Aleamoni, 1985; Swanson & Sisson, 1971).

Accuracy of Methods of Evaluation

There have also been charges that direct classroom observation, videotaping of classes, evaluation of course materials, and evaluation of instructor-graded student assignments are not accurate measures of a faculty member's teaching performance. Faculty members have often argued that evaluation rarely captures the essence of the act of teaching. McKeachie (1986) stated the thrust of this argument well, in noting that teaching "involves value judgments, and the means for achieving these values is complex. Research has revealed that many variables interact in determining faculty effectiveness" (p. 266).

Another scholar (Bulcock, 1984) expressed the belief that the accuracy of the evaluation can be compromised by focusing on superficial teaching skills (he called them "observables") rather than on more substantive (and often more complex) elements. Bulcock suggested that formative evaluation may be the most effective means of evaluating a fuller range of teaching abilities, because it is more likely that individual instructor styles can be taken into account in formative evaluation than in summative evaluation.

Other scholars noted how observers may affect an instructor's performance (Bergman, 1980; Britt, 1982; Gage, 1961; Hart, 1987; Sauter & Walker, 1976; Scriven, 1980; Stodolsky, 1984; Ward, Clark, & Harrison, 1981). In that regard, Gage (1961), in observing that some teachers may feel threatened by visitors, suggested that their performance may depend more on nerve than on skill.

Despite being a proponent of direct classroom observation, Hart (1987) also suggested that observers may affect the accuracy of an instructor's performance. Hart stated that "the very presence of the observer, however quiet and withdrawn, is an intervention that alters the situation, changes what is being observed. As one [pundit] put it, 'To observe a class is actually to observe a class being observed'" (p. 16).

To be considered valid measures, scholars and researchers have agreed, of course, that the methods employed must accurately evaluate teaching performance. Brandenburg, Braskamp, and Ory (1979), Centra (1975), and Stodolsky (1984) suggested that the process could be vastly improved by substantially increasing the number of visits and by evaluating all, or most, of the courses taught by instructors. The same scholars observed that these methods for improving the situation may be impractical because of the time required. To date, no satisfactory solutions to this dilemma have been offered.

Subjectivity

Faculty members have often charged that evaluation of teaching results in subjective, rather than objective, assessments of their performance. Several writers investigated this issue (Aleamoni, 1984; Arden, 1989; Bergman, 1980; Centra, 1986; Dressel, 1976; Edwards, 1974; Jones, 1986; McIntyre, 1978; Prater, 1983; Wood, 1977). Other scholars and researchers studied personality factors which may positively or negatively affect the accuracy of the

evaluations (Ballard, Reardron, & Nelson, 1976; Bulcock, 1984; Maslow & Zimmerman, 1956; Murray, 1975).

Bergman (1980), Centra (1975), and Jones (1986) cited the sociological research of Talcott Parsons (1954) as a basis for their observations that evaluation may become too subjective. These researchers noted that it is possible for evaluation to be based on ascription rather than achievement, on affectivity rather than neutrality, on diffuseness rather than specificity, on particularism rather than universalism, and on collectivity rather than self. In assessing the effects of these dichotomies on evaluation, Jones (1986) observed that the first factor in each pair represents a personal approach toward evaluation while the second factor represents a more bureaucratic approach. Bergman (1979, 1980), who was extremely critical of peer review, suggested that faculty members are too often inclined to employ the more personal of these approaches.

As this evidence has suggested, it is possible that the use of inappropriate criteria may result in nonobjective judgments. Yet, the use of inappropriate criteria seemed less likely in formative evaluation than in summative evaluation, because participants have less to lose (and much more to gain) when they, themselves, seek ways to improve the quality of their classroom teaching (Bulcock, 1984; Stodolsky, 1984; Sorcinelli, 1984). Still, the perception that evaluation might be based on subjective

impressions has been cited as a reason for the reluctance of faculty members to participate in various methods of formative evaluation.

Scriven (1980) succinctly summarized a number of objections to the use of direct classroom observation. These same objections may also be said to apply to videotaping of classes, evaluation of course materials, and evaluation of instructor-graded assignments. Scriven (1980) concluded:

> First, the visit itself alters the teaching, so that the visitor is not looking at a random sample. Second, the number of visits is too small to be an accurate sample from which to generalize, even if it were a random sample. Third, the visitor is not devoid of independent personal prejudices in favor of or against the teacher. (p. 10)

Enhancers

While scholars have noted a number of factors which may detract from the use of formative evaluation, they have also offered several suggestions for enhancing or improving the process. These recommendations have included the training of faculty in methods of supervision, the establishment of acceptable standards of teaching by the faculty, faculty interpretation of student ratings of courses and instructors, and the involvement of the faculty in the pre-planning of the programs.

Training in Methods of Supervision

Providing training to faculty members in methods of supervision was recommended by a number of writers in the field (Brock, 1981; Cancelli, 1987; Centra, 1975, 1986; Freer & Dawson, 1985; Heller, 1989; McIntyre, 1986; Mikula, 1979; Root, 1987; Seldin, 1984;

Sorcinelli, 1984). Cancelli (1987), Centra (1975), and McIntyre (1986) expressed the belief that inter-rater reliability of colleague evaluation could be improved if training in methods of supervision were made available to faculty members. In addition, Mikula (1979) suggested that this type of training might give faculty members insights into teaching and learning processes that they have not previously considered.

McIntyre (1986) developed a program for training faculty members to conduct evaluations of instruction. In this program, a number of faculty members visited the same classroom as a group, compared observations after the visit, and then attempted to arrive at a consensus about the strengths and weaknesses of the quality of the instruction. This method resulted in a considerable reduction in the variability of the assessment.

But researchers have also noted that training in methods of supervision does not eliminate all of the problems associated with improving inter-rater reliability. Centra (1986) observed that credible training programs require more time than many faculty members are willing to invest. Bergman (1980) concluded that "even with training, inappropriate criteria would still be quite influential in peer ratings--if only unconsciously" (p. 10).

While suggesting that teaching consultants are a good source for providing training in supervision, Sorcinelli (1984) acknowledged that it might also be provided by faculty members, because they have developed many of the skills required over the course of their

teaching careers. Agreeing, Brock (1981) suggested that whoever provides the training should have expertise in such areas as:

> audio-visual technology, ethnography, group dynamics, instructional evaluation, attribution theory, gaming and simulation, computer-assisted instruction, personalized systems of instruction, and philosophies of education. However, the attributes of greatest consequence for the consultant's effectiveness may be a commitment to student learning; an abiding curiosity about the relationship between teacher, student, and subject matter; an empathic disposition; a knowledge of local resources; a tendency toward self-disclosure; and effective interpersonal communication skills. (p. 239)

While Brock's listing seems imposing, it also demonstrated how complex the teaching and learning process is. Because of these complexities, scholars have begun to look at evaluation as an idiosyncratic and highly personal process (Bulcock, 1984; Cancelli, 1987).

Establishment of Standards of Effective Teaching by the Faculty

Because effective teaching is idiosyncratic and dependent on its context, scholars have concluded that it cannot be evaluated according to a single model of acceptable practice (Centra, 1986; Hart, 1987; Stodolsky, 1984). This conclusion has suggested that multiple standards be employed, with enough variety to accommodate differing teaching and learning styles and adaptable to a wide range of contexts.

Centra (1986) proposed a qualitative approach to evaluation, one that is consistent with the fundamental tenets of formative evaluation theory and with the conclusion that evaluation must be

adaptable to a variety of teaching and learning styles. In summarizing this view, Centra (1986) wrote:

> A qualitative approach would involve descriptions of classroom instruction based on the perceptions of the observers. . . [D]escriptions by several observers will more likely reflect possible personal biases and the resulting narrative could be much more useful [than] rating scales and numerical judgments. (pp. 3-4)

Centra was convinced that those using qualitative methods of evaluation result in looking at teaching in more depth than is possible through quantitative methods. The same writer also concluded that, like the quantitative approaches, qualitative methods require more time that many faculty members are willing to invest.

Interpretation of Student Ratings

Student evaluation of faculty teaching performance has been, and continues to be, employed more often to evaluate instruction than any other method (Seldin, 1984). Most writers in the field have acknowledged that students should be consulted in the process, but have generally agreed that students should not be the only source of information in evaluating instruction.

Many researchers have studied the relationship between student ratings and such variables as class size, expected course grade, time of day, required versus elective course, the subject matter, and so on. While there is some disagreement as to how these factors affect student ratings, a number of scholars suggested that these factors be taken into account when personnel decisions are made and when instructional improvement plans are formulated (Cohen & McKeachie, 1981; McKeachie, 1986).

Cohen and McKeachie (1981), Craig, Redfield, and Galluzzo (1986), and McKeachie (1986) recommended that the faculty play a role as integrators of information in comprehensive programs of faculty evaluation. Cohen and McKeachie (1981) recommended specifically that "student ratings should be evaluated by [faculty] peers who know the circumstances under which a particular course was taught" (p. 151). These scholars emphasized that student ratings cannot be taken at face value, as these ratings are not necessarily valid measures of performance effectiveness.

Faculty members themselves have been divided on the efficacy of peer review of student ratings as part of a more encompassing program of evaluation. For example, Britt (1982) found that only about 36% of faculty members believed that student ratings should be considered. In another study, Dienst (1981) found that faculty members supported peer review of student evaluations more than they supported direct classroom observation or evaluation of course materials, but only about 33% of them supported it, nevertheless.

Involvement of the Faculty in the Pre-Planning of Programs of Instructional Improvement

Nearly all writers in the field agreed that the development of successful programs of peer evaluation are dependent on the support of the faculty and of top-level administrators (Brock, 1981; Freer & Dawson, 1985; Heller, 1989; Razor, 1979; Seldin, 1984; Skoog, 1980; Soderberg, 1986). Without the support of the

faculty, Skoog (1980) observed that the process might be carried out perfunctorily, rather than with a genuine commitment to instructional improvement.

The essence of the argument for eliciting the support of the faculty in the development and implementation of programs of instructional improvement was well stated by Heller (1989):

> A decision to engage in peer supervision has to come from the peer group itself. The key is ownership; if teachers do not feel they own the project, then they will think somebody in the central office has a pet idea that is being forced on them. (p. 13)

Evaluation has often been seen as an adversarial relationship between those who are to be evaluated and those who will conduct the evaluation. Brock (1981) cautioned that careful attention:

> be given to the design of the procedures, to the inclusion of teachers in the process of design of the procedures, and especially to clear and repeated communication with teachers about the procedures. . . . With the reduction of threat comes the increased likelihood that teachers will effectively use evaluation data to make decisions about change in their teaching practices. (pp. 235-236)

The recommendations of Brock and Heller were incorporated into the program of formative evaluation described by Freer and Dawson (1985). This program, in which a reduction of the adversarial relationship was a primary goal, included seven recommendations for developing the program: (a) the commitment of adequate funding; (b) the involvement of as many teachers as possible in the initial planning stages; (c) an attempt to arrive at consensus when decisions are made, so that teachers "buy into" the program; (d) teacher involvement in training programs or courses in methods of

supervision; (e) collaboration among participants as the program is implemented; (f) involvement of teachers in the monitoring and fine-tuning of the program; and (g) a separation of the functions of summative and formative evaluation.

Freer and Dawson (1985) developed a system of formative evaluation that is consistent with the views of other scholars. According to Freer and Dawson, the program was received with more enthusiasm than the methods of evaluation that had been employed in the past. They were also convinced that the involvement of the faculty in the pre-planning was a key element in the program's success.

Personal and Institutional Benefits

Scholars have suggested that a number of benefits might come about as a result of faculty participation in formative evaluation. Five of these potential benefits were investigated in this study.

Improvement in the Quality of Instruction

Formative evaluation of instruction was conceived and implemented because its proponents believe that it is a viable means for improving the quality of teaching. Programs of formative evaluation involving colleagues as evaluators came about because a number of scholars are convinced that peers are more qualified than either students or administrators to assist the faculty in improving their teaching.

Few, if any, writers in the field have concluded that peers are the only source of pertinent information in the effort to improve

instruction or that peers are the best judges of all aspects of teaching. Rather, most writers have argued that peers are one of several sources of relevant information.

While a number of programs of formative evaluation have been implemented at colleges and universities where there are faculty development officials, other programs have been organized and developed by faculty members who are committed to instructional improvement. In these programs, most of the participants have expressed satisfaction with the results, believing that their teaching has improved because of the assistance they received from colleagues. While it is distinctly possible that instruction has improved as a result of participation, researchers noted that these programs have too often been evaluated by surveying the opinions of the participants rather than by conducting more sophisticated empirical studies (Erickson & Erickson, 1979; Hoyt & Howard, 1978; Levinson-Rose & Menges, 1981).

A few empirical studies have been conducted (Erickson & Erickson, 1979; Hoyt & Howard, 1978). In their studies, the researchers suggested that student ratings of faculty members have improved where faculty members have participated in the programs of formative evaluation. In the studies, the students rated their instructors both at midterm and near the end of the semester, with the earlier rating used as a covariate in the data analysis. Most of the results have shown statistically significant, though modest,

differences between program participants and nonparticipants, with participants showing the most improvement.

Still, researchers have emphasized that the evaluation of teaching improvement programs is far from refined. They have recommended that further study be conducted to evaluate larger numbers of participants, the effects of programs over longer periods of time, and the effects of volunteerism as opposed to a wider range of participants (Levinson-Rose & Menges, 1981).

In general, researchers have concluded that there is some basis for claims that formative evaluation is successful in helping teachers to improve. In addition, Erickson and Erickson (1979) indicated that faculty members who participated in the programs believed that it was "useful and well worth their time and effort, and that it results in significant, positive, and lasting changes in their classroom teaching skill performance" (p. 683).

Improvement in Student Learning

It has generally been assumed that the ultimate measure of success of the act of teaching is the maximization of student learning. By extrapolation, it has also been assumed that the ultimate goal of the formative evaluation of instruction is improved student learning. While recognizing that student learning is a primary focus of teaching and teacher improvement programs, scholars have emphasized that other factors also contribute to (or detract from) student learning. In that regard, Erickson and Erickson (1979) observed:

> It is difficult to deny the attractiveness of student learning gains as criteria for judging teaching improvement services, but we may have to defer their use as major criteria until more practical and powerful evaluation methodologies are available for dealing with the confounding influences of textbooks and peers. . . . (p. 671)

Bulcock (1984) identified a number of other potential intervening variables: (a) family background, (b) motivation, (c) aptitude, (d) course load, and (e) class size. This scholar observed and concluded:

> Efforts to measure teaching effectiveness on the basis of its impact on student learning is unpopular. This is because most teachers recognize that student learning is a multicausal activity, and that many significant factors . . . fall well outside the control of the teacher. Thus, to hold teachers responsible for the learning behaviors of their students is unreasonable. (p. 8)

Interest in improving student learning through formative evaluation has remained a goal for many faculty members. In a study of faculty attitudes about colleague evaluation of teaching, Britt (1982) found that 72% of respondents believed that teaching and learning would be improved through colleague evaluation of instruction. While the faculty may have a number of other agendas for participating in these programs, student achievement outcomes have apparently remained one of their chief concerns.

The Effects of Peer Review of Instruction on Faculty Morale and Collegiality

Reports on the effects of colleague evaluation of instruction on group morale and collegiality were mixed. Scholars generally agreed that peer review in the process of summative evaluation

affected faculty morale and collegiality adversely (Brandenburg, Braskamp, & Ory, 1979; Braskamp, 1978; Gunn, 1982; McIntyre, 1978; Sorcinelli, 1984). On the other hand, scholars generally found that the involvement of colleagues in the formative evaluation of instruction affected morale and collegiality positively (Cross, 1986; Edwards, 1974; Freer & Dawson, 1985; Heller, 1989; Menges, 1985; Roper, Deal, & Dornbusch, 1976; Shatzky & Silberman, 1986; Skoog, 1980; Sorcinelli, 1984).

When faculty members were asked if classroom observation would lower faculty morale, nearly 20% of the respondents indicated that it would, while almost 50% expressed the opposite view (Britt, 1982). However, interpreting this finding was difficult, since the researcher did not ask respondents to differentiate its effects in summative and formative evaluation.

Three writers in the field have commented on how faculty morale and collegiality have been affected by programs of formative evaluation of instruction. In the earliest of these studies, Skoog (1980) concluded that formative evaluation helps:

> faculty members acquire knowledge, insights, and strategies useful for self-supervision and self-improvement. Also, as a team works together, supportive relationships are established and discussions concerning teaching become more common, lengthy, and sophisticated. Ownership of common and unique teaching problems is acknowledged more openly. Increased satisfaction and pride in teaching can result. (p. 24)

In another report, Heller (1989) concluded that:

> using teachers in a peer supervision role is linked to their personal growth, their sense of collegiality, and to improved instructional practices--all of which

> contribute to higher morale, greater job satisfaction, improved school climate, and ultimately higher student achievement. (p. 11)

Galm (1985) noted that senior faculty members are often skeptical--or even cynical--about the value of programs of instructional improvement. This scholar reported on how the morale of senior faculty members was positively affected by a program of formative evaluation of instruction specifically designed for them. Galm was surprised by positive comments from participants, and noted that:

> working with this post-tenure study group, I got a completely different sense of my department, one that was in my bones but, because of the distractions of student complaints, not always in my head. I experienced the power of eight teachers and colleagues, a solid core of a permanent faculty of fifty, showing their concern for teaching and demonstrating their accumulated skill. It was awesome, and a fine corrective for any department chair who may have become cynical or have lost insight of the great faculty power in his or her department. (p. 67)

While Galm, Heller, and Skoog suggested that there is a positive relationship between participation in programs of formative evaluation conducted by colleagues and positive faculty morale and collegiality, they also indicated that the most successful programs have been implemented only after careful study and pre-planning. In most instances, the programs have been conducted with the support of the faculty (and administration), have involved the faculty in the initial planning and implementation stages, and have relied on voluntary participation. Nevertheless, conclusions about morale and collegiality have been reached after study of

small-scale programs. The effect on faculty morale and collegiality from more ambitious projects has not been studied.

Summary and Direction

The review of related literature and research has suggested to this researcher that there is a need for systematic study of peer involvement in formative evaluation of instruction. This conclusion was based on several factors: summative evaluation has not resulted in significant improvement in instruction; improvement in teaching is at least as important as, and perhaps more improvement than, benefits derived from summative evaluation; and, colleagues have insights into the processes of teaching and learning not possessed to the same degree by either students or administrators.

Even after a careful review of the literature, a number of questions concerning peer review in the formative evaluation of instruction remained unanswered. These questions included: Would faculty members participate in programs of formative evaluation if these programs were in place? Why are faculty members involved so infrequently in the process of the formative evaluation of their colleagues? Are there factors which detract from the process to such a degree that it is rejected by the faculty as a viable option to the types of evaluation conducted by students and administrators? Are there factors which would increase the likelihood that it would be accepted by (or made more acceptable to) the faculty? Of what value is formative evaluation to students, the faculty, and the institution? Answers to these questions were sought by surveying

faculty members at selected colleges and universities and by analyzing the results in light of what was learned through the review of literature.

CHAPTER III

METHODOLOGY FOR THE STUDY

A review of the literature on faculty evaluation has provided evidence that college teaching would be improved if programs of formative evaluation were in place. The research literature has also suggested that peer review would be a useful component in the process of formative evaluation of instruction. These claims have been substantiated by reports from program participants and by a limited number of credible experimental studies that demonstrate that modest, though statistically significant, improvements in student ratings of instructors have occurred as a result of taking part in programs of this type.

Yet, a number of scholars have observed that formative evaluation is seldom an integral part of programs of faculty evaluation in higher education. For example, in a pilot study that led to the present study, Keig (1989) reported that only 2 member institutions of the IAICU include formative evaluation as a key element in the process of faculty evaluation.

The procedures employed in this investigation have been outlined below under the following main headings: (a) Source of Data, (b) Selection of Sample, (c) Instrumentation, (d) Collection and Treatment of Data, and (e) Statistical Design.

Source of Data

The population for this study consisted of 2073 full-time faculty of the 26 baccalaureate degree-granting institutions

affiliated with the IAICU. Those individuals were identified in March and April, 1990, by examining faculty directories of the most recent (1988-90 or 1989-1991) catalog editions of the 26 colleges and universities. In this process, an attempt was made to identify only the men and women who held full-time appointments and to exclude those who taught part-time and/or held administrative appointments above the departmental (or divisional) level.

Selection of Sample

In order to assess the findings at the 95% confidence level, with a margin of error of plus or minus 5%, 324 faculty members from the population of 2074 were selected. The formula used in calculating the sample size has been included as Appendix A.

Three studies were used in estimating the response rate. In a pilot study of the questionnaire for this study, responses were received from exactly half of those who were asked to complete the survey and comment on the instrument itself. In another study of faculty attitudes toward peer review of instruction, questionnaires were returned by 55.4% of those who were invited to participate (Britt, 1982). In still another study of faculty attitudes toward faculty evaluation, responses were received from 54% of those who were asked to complete a questionnaire (Wilson, Dienst, & Watson, 1973).

Since it appeared that responses might be expected from approximately half of those asked to participate, it was initially decided that 650 questionnaires would be mailed to a random sample

of the population. When it was ultimately decided that the questionnaire would be mailed at the beginning of the 1990-91 academic year rather than at the end of the 1989-90 year, the number was increased to 750 to compensate for attrition could be expected among faculty members listed in the 1988-90 and 1989-91 catalogs used to identify the population and select the sample.

The names of the 751 faculty members (the process of rounding off yielded 751, rather than 750, names) invited to take part in the study were selected by randomly sampling a prorated number of faculty members from each institution. The number of questionnaires sent to each institution was based on the ratio of full-time faculty members at each of these colleges and universities to the population of 2073. The number of questionnaires mailed to each institution has been included as Appendix B.

Instrumentation

An original questionnaire was developed to answer four research questions and to test four hypotheses about methods of peer review in the formative evaluation of instruction and factors that were suggested in the literature as affecting the willingness of faculty members to participate in such programs. Included in the items selected were questions that surfaced following a pilot study of academic administrators affiliated with member institutions of the IAICU. A dissertation study of the attitudes of faculty members at three state universities toward summative and formative evaluation

involving colleagues as evaluators (Britt, 1982) also provided direction in determining which items to include in the questionnaire.

The main body of the questionnaire consisted of 29 content items in Yes-No-Not Sure and Likert-type formats. Those items were arranged in seven groups. Respondents were asked to indicate: (a) which, if any, of four methods of formative evaluation of instruction they would participate in if that evaluation were performed by informed peers (four items); (b) what factors, if any, might detract from each of these four methods to such a degree that the methods might be considered unacceptable (four items in each of four groups, for a total of 16 items); (c) what factors, if any, might enhance the process or increase the likelihood that faculty members might find these types of evaluation acceptable (four items); and (d) how students, faculty, and/or the institution might benefit by having such programs in place (five items).

Eight demographic items in multiple-choice and completion-type formats were also included. In this section of the questionnaire, respondents were asked to indicate their: (a) gender, (b) age, (c) years of teaching experience in postsecondary education, (d) academic rank, (e) tenure status, (f) institution's academic calendar arrangement, (g) average teaching load during the equivalent of a semester of instruction, and (h) academic discipline. The letter of transmittal and the questionnaire have been included as Appendixes C and D.

On April 11, 1990, a preliminary draft of the questionnaire was distributed via campus mail to 20 faculty members in the College of Education at the University of Northern Iowa. In an accompanying cover letter, these faculty members were asked to complete the questionnaire and to offer suggestions for strengthening it. Ten questionnaires (50%) were returned. As a result of the comments made by faculty members, some questionnaire items were clarified and one group of items was deleted from the final version of the questionnaire.

Prior to the final printing, the questionnaire and letter of transmittal were submitted to the Human Subjects Review Board of the University of Northern Iowa for approval. Permission to proceed with the study was granted on September 18, 1990.

Collection and Treatment of Data

On October 9, 1990, a questionnaire, a letter of transmittal, and a postage-paid return envelope were mailed to 751 potential respondents via pre-sorted, first-class mail. From October 11, 1990, through December 14, 1990, 372 responses (49.5%) were returned.

As the questionnaires were returned, the researcher opened the envelopes, examined the contents, numbered the questionnaires according to the date of receipt, and coded the academic discipline of the respondent. Computer data entry began on October 29, 1990, and continued through December 14, 1990. Computer data entry verification was carried out from December 17-20, 1990.

Statistical Design

While descriptive statistics were calculated for each questionnaire item, relationships among certain variables were the principal focus of this study. Determining what relationships to explore was based on Research Questions 2, 3, and 4 and the predicted relationships in Hypotheses 2, 3, and 4.

The Statistical Package for the Social Sciences (SPSS) was used for preparing the data for analysis and for providing statistical treatment of the data. The SPSS program was prepared at the Center for Social and Behavioral Research (CSBR) at the University of Northern Iowa.

Descriptive Statistics

For each Yes-No-Not Sure content item on the questionnaire, frequency counts and percentages were calculated. On all Likert-type content items, means and standard deviations, as well as frequency counts and percentages, were calculated.

For the demographic items, frequency counts and percentages were calculated on dichotomous and categorical variables. These items were those in which respondents were asked to indicate their gender, academic rank, tenure status, and institution's type of academic calendar. Means, standard deviations, and medians, as well as frequency counts and percentages, were calculated for demographic items on interval scales. Those items were the respondent's age, number of years of teaching experience at the

college or university level, and number of credit hours taught during a typical semester or quarter of instruction.

On the demographic item in which respondents were asked to list the discipline in which they teach all or most of their courses, each discipline listed by respondents was assigned a code number. At the same time, each discipline was assigned to one of several categories of disciplines for possible future data analysis. The codes for each discipline and the categories of disciplines have been included as Appendix E.

Four research questions were studied and four corresponding hypotheses were tested. Three of the hypotheses were tested at the .05 level of significance, using non-directional tests. Answering the first research question and testing its corresponding hypothesis required only the reporting of descriptive statistics. This research question and hypothesis have been stated as follows:

Research Question 1. What proportion of faculty members would avail themselves of direct classroom observation, videotaping of classes, evaluation of course materials, and evaluation of instructor-graded student assignments if these methods were available for the purpose of formative evaluation?

Hypothesis 1: A majority of faculty members would avail themselves of direct classroom observation, videotaping of classes, evaluation of course materials, and instructor-graded student assignments if these methods were in place in formative evaluation involving colleagues as evaluators.

Correlational Procedures

Answering the remaining research questions and testing the hypotheses corresponding to each of the research questions required correlational procedures. Research questions and hypotheses of greatest interest to the researcher were those for which relationships among variables were investigated.

The second research question and corresponding hypothesis pertained to relationships among faculty members who would take part in each of four methods of formative evaluation, and four corresponding factors for each method that have been suggested by researchers as detracting from the likelihood that faculty members would avail themselves of each of these methods. The research question and corresponding hypothesis have been stated as follows:

Research Question 2. What is the relationship between each of the methods of evaluation and each of its corresponding detractors?

Hypothesis 2: There will be a negative relationship between attitudes toward each method of evaluation and each of its corresponding detractors.

For each of the four methods of evaluation, respondents were asked to indicate their degree of agreement (or disagreement) with the following statements:

Direct classroom observation (or videotaping of classes, evaluation of course materials, evaluation of instructor-graded

student assignments, as appropriate) would . . .

* infringe on my academic freedom.

* probably not measure accurately my teaching performance.

* probably not be representative of a typical teaching-learning situation.

* result in a subjective, rather than an objective, assessment of my performance.

Another thrust of the study was to investigate the relationships among faculty members who would avail themselves of the four methods of evaluation and four factors that have been suggested as enhancing the process of formative evaluation. The research question and corresponding hypothesis have been stated as follows:

Research Question 3. What is the relationship between each of the methods of evaluation and each of the formative evaluation enhancers?

Hypothesis 3: There will be a positive relationship between attitudes toward each method of evaluation and each of the formative evaluation enhancers.

For this question and hypothesis, respondents were asked to indicate their degree of agreement (or disagreement) with the following statements:

A program of instructional evaluation in which informed peers evaluate the teaching performance of colleagues would be enhanced if . . .

* training in evaluation methods was made available to faculty.

* acceptable standards of teaching were established by the faculty.

* student ratings of participants were included.

* parties involved in the process were consulted in the pre-planning.

The fourth research question and corresponding hypothesis pertained to relationships among respondents who would take part in each of the four methods of evaluation, and their views toward the potential value of peer review on students, faculty, and/or the institution in the process of instructional evaluation. The research question and hypothesis have been stated as follows:

Research Question 4. What is the relationship between each method of evaluation and each of the individual/institutional improvement factors?

Hypothesis 4: There will be a positive relationship between attitudes toward each method of evaluation and each of the individual/institutional improvement factors.

For this question and hypothesis, respondents were asked to indicate their degree of agreement (or disagreement) with the following statements:

A program of instructional evaluation in which informed peers evaluate the teaching performance of colleagues would improve . . .

* the quality of instruction.
* student learning.
* the tenure status of junior faculty members.
* the morale of senior faculty members.
* the collegial climate of the college/university.

Summary

From the population of 2074 full-time faculty members of the 26 baccalaureate degree-granting member institutions of the IAICU, a random sample of 751 faculty members was invited to participate in this study. To ensure that each institution would be proportionately represented in this sample, a random sample from each college and university was drawn in the ratio of its number of full-time faculty members to the population of 2073. The questionnaire that faculty members were asked to complete consisted of 29 content items in Yes-No-Not Sure and multiple-choice formats and eight demographic items in multiple-choice and completion-type formats. Questionnaires were received from 372 respondents.

The statistical procedures employed for this study were designed to provide answers to four research questions and to test hypotheses corresponding to each of these questions. Descriptive statistics, correlational procedures, and tests of statistical significance were employed. Data analyses based on these statistical procedures have been selected as the focus for the next chapter of this study.

CHAPTER IV

PRESENTATION AND ANALYSIS OF THE DATA

The purpose of this study was to examine faculty attitudes toward methods of peer review in the formative evaluation of instruction in higher education, and toward selected variables that may affect the willingness of faculty members to avail themselves of these methods. Those selected variables were what have been called detractors, enhancers, and improvement factors.

A random sample of 751 faculty from the baccalaureate degree-granting institutions affiliated with the IAICU was requested to respond to a questionnaire about the methods of evaluation and the other variables. Presented in this chapter are the results of this investigation.

Response to the Questionnaire

From October 11 through December 14, 1990, 372 respondents returned usable questionnaires. This number represented a response rate of approximately 49.5%. The return of 372 usable questionnaires exceeded by 48 the number required to meet the 95% confidence level for reliability.

Demographic Breakdown of Responses

Besides being asked to respond to several content items, respondents were asked to provide demographic information about themselves and about the institution at which they teach. Of the respondents, 253 (68.4%) were men, and 117 (31.6%) were women. The mean age of respondents was 46.6 years (sd = 9.7); the ages

ranged from 24 to 69 years, with a median age of 46 years. The mean number of years of teaching experience at the college or university level was 15.4 years (sd = 9.4); teaching experience ranged from less than one year to 40 years, with a median of 15 years. Of those responding, 137 (36.8%) held the rank of professor, 95 (25.5%) were associate professors, 100 (26.9%) were assistant professors, and 30 (8.1%) were instructors (or the equivalent). Ten (2.7%) held other ranks, including those with part-time or adjunct status. Of those completing the questionnaire, 226 (60.9%) were tenured; 145 (39.1%) were non-tenured.

Respondents listed 54 different disciplines in which they taught all or most of their courses. When these disciplines were grouped into categories, 121 faculty members (32.6%) taught in the arts and humanities; 20 (5.4%) in health care; 82 (22.0%) in mathematics and the sciences; 57 (15.4%) in the social sciences; 24 (6.4%) in health, physical education, and recreation; 29 (7.8%) in education; and 37 (10.0%) in business.

Answering the Research Questions and Testing the Hypotheses

Four research questions were posed and a hypothesis corresponding to each of these questions was tested. Hypotheses were tested at the .05 level of significance using non-directional tests. In calculating the Pearson correlation coefficients, pairwise deletion of data was employed.

The First Research Question and Its Corresponding Hypothesis

Research Question 1 was asked and Hypothesis 1 was tested to determine in which, if any, of four methods of formative evaluation of instruction faculty members would participate if these methods were conducted by peers. The hypothesis was stated as follows: A majority of faculty members would avail themselves of direct classroom observation, videotaping of classes, evaluation of course materials, and evaluation of instructor-graded student assignments if these methods were in place in formative evaluation involving colleagues as evaluators.

As indicated in Table 1, a majority of faculty members reported that they would avail themselves of all four methods of peer review when used in a process of formative evaluation. Direct classroom observation and evaluation of course materials were supported overwhelmingly (93.5% and 94.3%, respectively). Videotaping of classes and evaluation of instructor-graded student assignments (at 61.9% and 77.9%, respectively) were also strongly supported. Thus, all components of Hypothesis 1 were supported.

The statistics do not indicate why there is more support for direct classroom observation and evaluation of course materials than for videotaping of classes and evaluation of instructor-graded student assignments. Direct classroom observation may be supported because more has been written about it and because it has been employed more often in practice (albeit much more so in elementary

Table 1

Number and Percentage of Respondents Who Would Participate in Four Methods of Peer Review in the Formative Evaluation of Instruction

Method	Yes		No		Not Sure	
	N	Percent	N	Percent	N	Percent
Direct Classroom Observation	346	93.5	11	3.0	13	3.5
Videotaping of Classes	227	61.9	63	17.2	77	21.0
Evaluation of Course Materials	348	94.3	9	2.4	12	3.3
Evaluation of Graded Student Assignments	285	77.9	43	11.7	38	10.4

and secondary schools than in colleges and universities) than the other methods. Evaluation of course materials may be perceived as personally less threatening to faculty members than any of the other methods. Nevertheless, support for videotaping was strong. Support for evaluation of instructor-graded student assignments, a method rarely employed at any level of schooling, was even stronger than that for videotaping.

The Second Research Question and Its Corresponding Hypothesis

The second of the research questions and hypotheses pertained to relationships among faculty members' willingness to take part in the four methods of formative evaluation, and their attitudes toward factors that have been suggested by scholars as detracting from faculty participation in each of the methods. The hypothesis was stated as follows: There will be a negative relationship between attitudes toward each method of evaluation and each of its corresponding detractors.

Direct classroom observation. As predicted, negative relationships were found between faculty members who would participate in direct classroom observation and each corresponding detractor. The findings shown in Table 2 suggest that faculty members who say they would participate in this form of evaluation did not view the detractors in an adverse manner. Thus, all components of Hypothesis 2 with respect to direct classroom observation were supported.

Videotaping of classes. As hypothesized, negative relationships were found between videotaping of classes and each of its detractors. Again, the findings suggest that faculty members who would participate in this form of evaluation do not view the detractors in an adverse manner. Thus, all components of Hypothesis 2 with respect to videotaping of classes were supported.

Table 2

Relationships Between Faculty Members' Willingness to Participate in Direct Classroom Observation and Their Agreement With Each Detractor

	Direct Classroom Observation		
Detractor	Correlation Coefficient	Number	Level of Significance
Infringe on Academic Freedom	-.404	367	.001
Not Measure Accurately Teaching Performance	-.253	366	.001
Not Represent Typical Learning Situation	-.280	368	.001
Subjective, Rather Than Objective, Assessment	-.216	365	.001

As indicated in Table 3, the Pearson correlation coefficients ranged from low to moderate, according to Cohen's (1977) conventions. Relationships between direct classroom observation and videotaping of classes and all but one detractor are similar.

Evaluation of course materials. As predicted, negative relationships were found between faculty members who would participate in evaluation of course materials and their agreement

Table 3

Relationships Between Faculty Members' Willingness to Participate in Videotaping of Classes and Their Agreement With Each Detractor

	Videotaping of Classes		
Detractor	Correlation Coefficient	Number	Level of Significance
Infringe on Academic Freedom	-.388	362	.001
Not Measure Accurately Teaching Performance	-.418	364	.001
Not Represent Typical Learning Situation	-.339	363	.001
Subjective, Rather Than Objective, Assessment	-.235	362	.001

with each detractor. As the significance levels in Table 4 infer, all components of Hypothesis 2 with respect to this method of evaluation were supported. These statistics suggest that faculty members who would participate in evaluation of course materials do not view the detractors in an adverse manner.

Evaluation of instructor-graded student assignments. The correlation coefficients between faculty members' willingness to participate in instructor-graded student assignments and their

Table 4

Relationships Between Faculty Members' Willingness to Participate in Evaluation of Course Materials and Their Agreement With Each Detractor

	Evaluation of Course Materials		
Detractor	Correlation Coefficient	Number	Level of Significance
Infringe on Academic Freedom	-.277	366	.001
Not Measure Accurately Teaching Performance	-.209	366	.001
Not Represent Typical Learning Situation	-.213	366	.001
Subjective, Rather Than Objective, Assessment	-.161	363	.002

agreement with each of its detractors are, in all instances, moderate. The results presented in Table 5 show that all components of Hypothesis 2 with respect to evaluation of instructor-graded student assignments were supported.

Summary. While all components of Hypothesis 2 were supported, the relationships among faculty members who would participate in these methods of formative evaluation and their attitudes toward

Table 5

Relationships Between Faculty Members' Willingness to Participate in Evaluation of Instructor-Graded Student Assignments and Their Agreement With Each Detractor

	Evaluation of Graded Student Work		
Detractor	Correlation Coefficient	Number	Level of Significance
Infringe on Academic Freedom	-.534	361	.001
Not Measure Accurately Teaching Performance	-.416	361	.001
Not Represent Typical Learning Situation	-.340	360	.001
Subjective, Rather Than Objective, Assessment	-.304	357	.001

corresponding detractors ranged from moderate to small. That fact notwithstanding, faculty members have not agreed, in general, with the conclusions of writers in the field that faculty members are averse to participating in these methods of formative evaluation. Apparently, they do not believe that their academic freedom will be threatened, and are not particularly concerned about the accuracy,

the representativeness, or the degree of objectivity (or subjectivity) of these methods of evaluation.

The Third Research Question and Its Corresponding Hypothesis

The third research question and its hypothesis pertained to relationships between faculty members who would take part in the same methods of evaluation and each of four factors that have been suggested by scholars as enhancing the process of formative evaluation. The four factors investigated were the training of faculty members in methods of evaluation, the establishment of acceptable standards of teaching by the faculty, the examination of student ratings of courses and instructors by fellow faculty members, and the involvement of the faculty in the pre-planning process. The third hypothesis was stated as follows: There will be a positive relationship between attitudes toward each method of evaluation and each of the formative evaluation enhancers.

Table 6 was prepared to illustrate the degree of support for each of the enhancers. These data show widespread support for each enhancer. As Tables 7 through 10 show, however, very small correlations were found among the methods of evaluation and the enhancers.

Direct classroom observation. As the data displayed in Table 7 infer, the relationships between faculty members who would take part in direct classroom observation and their agreement with each enhancer were not significant. Despite substantial support for

Table 6

Degree of Support of Faculty Members for the Formative Evaluation Enhancers

Enhancer	Degree of Agreement SA	A	U	D	SD
Training of Faculty in Evaluation Methods					
Frequency	133	188	26	13	6
Percent	36.3	51.4	7.0	3.5	1.6
Establishment of Standards of Teaching by Faculty					
Frequency	102	149	66	34	16
Percent	27.8	40.6	18.0	9.3	4.4
Examination of Student Ratings by Faculty Peers					
Frequency	54	174	75	47	17
Percent	14.7	47.4	20.4	12.8	4.6

(table continues)

Enhancer	Degree of Agreement SA	A	U	D	SD
Involvement of Faculty in Pre-planning of Process					
Frequency	145	186	26	8	1
Percent	39.6	50.0	7.1	2.2	.3

Note. Likert-type designators: S = Strongly Agree, A = Agree, U = Undecided, D = Disagree, SD = Strongly Disagree.

each of the enhancers, virtually no correlation was found between this method of evaluation and any of the enhancers. Therefore, none of the components of Hypothesis 3 with respect to direct classroom observation were supported.

Videotaping of classes. As Table 8 shows, the relationships between faculty members who would participate in videotaping of classes and their agreement with each enhancer are, according to Cohen's (1977) conventions, small. However, the relationship between faculty members who would participate in videotaping of classes and who support the involvement of the faculty in the pre-planning of programs of formative evaluation is statistically significant at the .008 level. The relationship between faculty members who would take part in videotaping and who support faculty review of

Table 7

Relationships Between Faculty Members' Willingness to Participate in Direct Classroom Observation and Their Agreement With Each Enhancer

	Direct Classroom Observation		
Enhancer	Correlation Coefficient	Number	Level of Significance
Training of the Faculty in Evaluation Methods	.011	366	.832
Standards of Teaching Established by the Faculty	.027	367	.638
Faculty Examination of Student Ratings	.028	367	.595
Involving the Faculty in the Pre-planning	.009	366	.871

student ratings of faculty performance (.053) is near the accepted value of .05. The relationships between videotaping and training the faculty in methods of evaluation and between videotaping and the setting of acceptable standards of teaching by the faculty are clearly nonsignificant. Therefore, the hypothesized relationship between videotaping and faculty involvement in pre-planning was

Table 8

Relationships Between Faculty Members' Willingness to Participate in Videotaping of Classes and Their Agreement With Each Enhancer

	Videotaping of Classes		
Enhancer	Correlation Coefficient	Number	Level of Significance
Training of the Faculty in Evaluation Methods	.052	363	.322
Standards of Teaching Established by the Faculty	.018	364	.735
Faculty Examination of Student Ratings	.102	364	.053
Involving the Faculty in the Pre-planning	.140	363	.008

supported, while the other components of the hypothesis with respect to videotaping were not supported. Yet, the finding of a statistically significant relationship between videotaping and faculty involvement in the pre-planning of programs of formative evaluation may have limited practical significance since the correlation coefficient between the two variables is small.

Evaluation of course materials. As Table 9 shows, the relationships between faculty members who would participate in peer evaluation of course materials in the process of formative evaluation of instruction and their agreement with each enhancer are small. Since none of the levels of significance are less than .05, none of the components of Hypothesis 3 with respect to course materials were supported.

Evaluation of instructor-graded student assignments. As Table 10 shows, the relationships between faculty members who would participate in evaluation of instructor-graded student assignments and their agreement with each enhancer are much the same as relationships between evaluation of course materials and the same enhancers. In all cases, the correlation coefficients are extremely small. Thus, none of the components of this hypothesis with respect to evaluation of instructor-graded student assignments were supported.

Summary. Except for the relationship between faculty members who would take part in videotaping of classes and who believe that the process of formative evaluation would be enhanced if the faculty were involved in the pre-planning of the programs, none of the hypothesized relationships among methods of evaluation and the formative evaluation enhancers are statistically significant. In every case, the correlation coefficients are small.

These findings should not be construed to indicate that faculty members do not find the enhancers important, as the enhancers are

Table 9

Relationships Between Faculty Members' Willingness to Participate in Evaluation of Course Materials and Their Agreement With Each Enhancer

	Evaluation of Course Materials		
Enhancer	Correlation Coefficient	Number	Level of Significance
Training of the Faculty in Evaluation Methods	.060	365	.445
Standards of Teaching Established by the Faculty	-.074	366	.156
Faculty Examination of Student Ratings	.045	366	.393
Involving the Faculty in the Pre-planning	-.020	365	.709

supported overwhelmingly by respondents. It should be concluded only that, with one exception, the hypothesized relationships were not found.

Table 10

Relationships Between Faculty Members' Willingness to Participate in Evaluation of Instructor-Graded Student Assignments and Their Agreement With Each Enhancer

	Evaluation of Graded Student Work		
Enhancer	Correlation Coefficient	Number	Level of Significance
Training of the Faculty in Evaluation Methods	.055	363	.293
Standards of Teaching Established by the Faculty	.005	364	.930
Faculty Examination of Student Ratings	.004	363	.938
Involving the Faculty in the Pre-planning	.034	362	.521

The Fourth Research Question and Its Corresponding Hypothesis

The fourth research question and its corresponding hypothesis pertained to the relationships between faculty members who would participate in the four methods of evaluation and their attitudes toward each of five individual/institutional improvement factors.

The improvement factors investigated were quality of instruction, student learning, the tenure status of junior faculty members, the morale of senior faculty members, and the collegial climate of the college or university. The fourth hypothesis was stated as follows: There will be a positive relationship between attitudes toward each method of evaluation and each of the individual/institutional improvement factors.

Table 11 was prepared to show the degree of support of faculty members for each of the improvement factors. The data show that sizeable majorities of respondents believe that the quality of teaching, student learning, and the tenure success of junior faculty members would be improved as a result of faculty participation in formative evaluation. However, minorities of respondents expressed the view that the morale of the senior faculty and the collegial climate of the college or university would be improved by faculty participation in formative evaluation.

Direct classroom evaluation. As shown in Table 12, the correlation coefficients between faculty members who would participate in direct classroom observation and their attitudes toward the benefits that might be derived from peer involvement in the process of formative evaluation of instruction are, according to Cohen's (1977) conventions, small. Nevertheless, the relationship between direct classroom observation and improvement in the quality of instruction is statistically significant. Therefore, that component of Hypothesis 4 was supported.

Table 11

Degree of Support of Faculty Members for Personal and Institutional Benefits That Might Occur as a Result of Peer Involvement in Instructional Evaluation

	Degree of Agreement				
Improvement	SA	A	U	D	SD
Quality of Instruction					
Frequency	62	220	69	17	1
Percent	16.8	59.6	18.7	4.6	.3
Student Learning					
Frequency	47	164	121	33	2
Percent	12.8	44.7	33.0	9.0	.5
Tenure Status of the Junior Faculty					
Frequency	41	180	119	23	3
Percent	11.2	49.2	32.5	6.3	.8
Morale of the Senior Faculty					
Frequency	19	67	177	85	19
Percent	5.2	18.3	48.2	23.2	5.2
Collegial Climate of the Institution					
Frequency	36	121	138	59	13
Percent	9.8	33.0	37.6	16.1	3.5

Note. Likert-type designators: S = Strongly Agree, A = Agree, U = Undecided, D = Disagree, SD = Strongly Disagree.

Table 12

Relationships Between Faculty Members' Willingness to Participate in Direct Classroom Observation and Their Agreement With Each Improvement Factor

	Direct Classroom Observation		
Improvement Factor	Correlation Coefficient	Number	Level of Significance
Quality of Instruction	.160	369	.002
Student Learning	.091	367	.082
Tenure Status of the Junior Faculty	.056	366	.285
Morale of the Senior Faculty	.061	367	.243
Collegial Climate of the Institution	.090	367	.084

None of the other hypothesized relationships are statistically significant. Therefore, the hypothesized relationships between direct classroom observation and student learning, tenure status of the junior faculty, morale of the senior faculty, and collegial climate of the institution were not supported.

Videotaping of classes. Although the correlation coefficients between faculty members who would participate in videotaping of classes and their agreement with each of the individual/institutional improvement factors are relatively small, all are statistically significant, as Table 13 shows. Therefore, all components of Hypothesis 4 with respect to videotaping of classes were supported.

Table 13

Relationships Between Faculty Members' Willingness to Participate in Videotaping of Classes and Their Agreement With Each Improvement Factor

	Videotaping of Classes		
Improvement Factor	Correlation Coefficient	Number	Level of Significance
Quality of Instruction	.259	366	.001
Student Learning	.245	364	.001
Tenure Status of the Junior Faculty	.162	363	.002
Morale of the Senior Faculty	.199	364	.002
Collegial Climate of the Institution	.225	364	.001

In general, faculty members who would take part in videotaping of their classes in a process of formative evaluation also expressed the belief that students, faculty members, and colleges and universities will benefit as a result of having programs of this type in place.

Evaluation of course materials. As shown in Table 14, the relationships between faculty members who would participate in evaluation of course materials and their agreement with each improvement factor are small. However, the relationships between evaluation of course materials and quality of instruction and between evaluation of course materials and student learning are statistically significant. Therefore, components of Hypothesis 4 with respect to course materials and these two improvement factors were supported.

The relationships between course materials and the tenure status of the junior faculty, the morale of the senior faculty, and the collegial climate of the institution are not significant. Consequently, components of this hypothesis with respect to evaluation of course materials and these three variables were not supported.

Evaluation of instructor-graded student assignments. As Table 15 indicates, the correlations between faculty members who would participate in evaluation of instructor-graded student assignments and each improvement factors are small. However, the correlation

Table 14

Relationships Between Faculty Members' Willingness to Participate in Evaluation of Course Materials and Their Agreement With Each Improvement Factor

	Evaluation of Course Materials		
Improvement Factor	Correlation Coefficient	Number	Level of Significance
Quality of Instruction	.135	368	.009
Student Learning	.143	366	.006
Tenure Status of the Junior Faculty	.019	365	.724
Morale of the Senior Faculty	.041	366	.342
Collegial Climate of the Institution	.064	366	.221

between instructor-graded student assignments and improvement in student learning is statistically significant. Therefore, this component of Hypothesis 4 was supported.

The remaining correlation coefficients with respect to evaluation of instructor-graded student assignments are not

statistically significant. Consequently, the components of the hypothesis pertaining to evaluation of instructor-graded student assignments and quality of instruction, tenure status of the junior faculty, morale of the senior faculty, and collegial climate of the institution were not supported.

Table 15

Relationships Between Faculty Members' Willingness to Participate in Evaluation of Instructor-Graded Student Assignments and Their Agreement With Each Improvement Factor

	Evaluation of Graded Student Work		
Improvement Factor	Correlation Coefficient	Number	Level of Significance
Quality of Instruction	.102	365	.052
Student Learning	.162	363	.002
Tenure Status of the Junior Faculty	.084	362	.110
Morale of the Senior Faculty	.079	363	.134
Collegial Climate of the Institution	.097	363	.066

Summary. None of the correlation coefficients among faculty members who would participate in the methods of evaluation investigated in this study and the improvement factors are large (or, in most cases, even moderate). Yet, some of them are statistically significant. In general, the relationships among methods and improvement in the quality of instruction and improvement in student learning are the strongest. The weakest correlations are those with respect to improvements in the tenure status of the junior faculty and in the morale of the senior faculty. In all cases, lying somewhere between these two extremes are relationships among methods of evaluation and improvement in the collegial climate of the college or university.

Faculty members who would take part in videotaping of their classes were most likely to believe that peer involvement in instructional evaluation would improve the quality of instruction, student learning, the tenure status of the junior faculty, the morale of the senior faculty, and the collegial climate of the college or university. These findings suggest that faculty members who would participate in videotaping are likely to believe that peer involvement in instructional evaluation will lead to benefits to students, the faculty, and the college or university.

However, the findings also suggest that faculty members remain skeptical about improvements in the tenure status of the junior faculty, the morale of the senior faculty, and the collegial climate of the institution that scholars believe will result with faculty

involvement in the formative evaluation of instruction. Faculty members were more in agreement with scholars with respect to improvements in the quality of instruction and in student learning.

Other Findings

The hypothesized relationships among faculty members' willingness to participate in four methods of peer review in the process of formative evaluation and their agreement with a number of factors which may affect their willingness to participate were, of course, the principal foci of the data analysis. However, demographic breakdowns of some of the data were also of interest to the researcher.

Demographic breakdowns of respondents' willingness to participate in direct classroom observation, videotaping of classes, evaluation of course materials, and evaluation of instructor-graded student assignments were calculated. Chi square tests for independence between each method of evaluation and six demographic factors--gender, age (by categories), years of teaching at the college or university level (by categories), academic rank, tenure status, and academic discipline--did not result in any significant differences.

Overview and Direction

The results of this study revealed that: (a) there were statistically-significant negative relationships between faculty members' willingness to participate in methods of formative evaluation and their agreement with each of the enhancers;

(b) with one exception, there were no statistically-significant positive relationships between faculty members willingness to participate and their agreement with each of the enhancers; and (c) in general, there were statistically-significant positive relationships between faculty members' willingness to participate and their agreement that the quality of instruction and student learning would be improved by faculty participation in programs of formative evaluation. Three topics will be discussed in the final chapter: a summary of the findings, recommendations regarding the practice of peer review in the formative evaluation of instruction in higher education, and recommendations for further study.

CHAPTER V

SUMMARY AND RECOMMENDATIONS

The purpose of this study was to examine faculty attitudes toward methods of peer review in the formative evaluation of instruction in higher education, and toward selected variables that may affect the willingness of faculty members to avail themselves of these methods. The four methods investigated were direct classroom observation, videotaping of classes, evaluation of course materials, and evaluation of instructor-graded student assignments. The other factors examined pertained to what might detract from or enhance the process of formative evaluation of instruction, and to how students, the faculty, and the college or university might benefit from faculty participation in programs of formative evaluation.

Summary

This summary of the results of the study has been organized into two sections. These sections have been written to correspond to the procedures and findings presented in Chapters III and IV.

Methodology

From the population of 2074 full-time faculty members at institutions affiliated with the IAICU, a random sample of 751 faculty members was asked to participate in the study. From that sample, 372 usable questionnaires were returned by the December 14, 1990, cut-off date.

An original, 37-item questionnaire was designed to elicit answers to questions which were not answered in the review of related literature and research. These questions were the following: Would faculty members participate in programs of formative evaluation if such programs were in place? Why are programs of formative evaluation found so rarely at colleges and universities? Do factors identified by scholars as detracting from the process affect faculty members' willingness to participate? Do factors identified by scholars as enhancing the process affect faculty members' willingness to participate? Do factors suggested by scholars as benefitting students, the faculty, and/or the institution affect faculty members' willingness to participate? The questionnaire was made up of 29 content and eight demographic items.

As questionnaires were returned, they were coded for the disciplines in which the respondents taught all or most of their courses. The data were entered into the computer between October 29 and December 14, 1990. Verification of data entry took place between December 17 and December 20, 1990. The SPSS program for displaying the data was written on December 20, 1990.

Appropriate descriptive statistics were calculated for each item on the questionnaire. Correlation coefficients and tests of statistical significance were calculated on each component of Hypotheses 2, 3, and 4. These correlational procedures were carried out between faculty members' willingness to participate in each of the four methods of formative evaluation and their agreement with

each of four corresponding detractors, four formative evaluation enhancers, and five individual/institutional improvement factors.

Presentation and Analysis of Data

Data analysis revealed that the proportions of faculty members who would participate in each of the four methods of formative evaluation were large enough to support all components of Hypothesis 1. The data analysis also revealed that all 16 components of Hypothesis 2, one of 16 components of Hypothesis 3, and nine of 16 components of Hypothesis 4 were statistically significant.

Hypothesis 1. Hypothesis 1 was stated as follows: A majority of faculty members would avail themselves of direct classroom observation, videotaping of classes, evaluation of course materials, and evaluation of instructor-graded student assignments if these methods were in place in formative evaluation involving colleagues as evaluators. Large majorities of respondents said they would avail themselves of each method of formative evaluation if these methods were available.

Britt (1982) asked respondents which methods, if any, they found acceptable in the evaluation of instruction, without specifying summative or formative evaluation. The percentages of positive responses were: direct classroom observation (72.5%), videotaping of classes (26.8%), and evaluation of course materials (51.6%). These percentages are much lower than the positive responses of participants in the present study who said they would take part in

the same methods if the evaluation were performed by peers for the purpose of formative evaluation.

Hypothesis 2. Hypothesis 2 was stated as follows: There will be a negative relationship between attitudes toward each method of evaluation and each of its corresponding detractors. In all cases, the significance levels were smaller than the accepted level. Thus, all components of the hypothesis were accepted.

Scholars have speculated that faculty members may not be willing to participate in specific methods of formative evaluation because there are factors which detract from the value of these methods (Bulcock, 1984; Edwards, 1974; Scriven, 1980). In this study, contrary to what scholars speculated, an inverse relationship was found between faculty members' willingness to participate and their agreement that the detractors adversely affect willingness to participate.

Hypothesis 3. Hypothesis 3 was stated as follows: There will be a positive relationship between attitudes toward each method of evaluation and each of the formative evaluation enhancers. Only one component of this hypothesis was statistically significant, namely, the relationship between faculty members who would participate in videotaping of classes and who believe that the process would be improved by involving the faculty in the pre-planning of formative evaluation. While none of the other relationships were statistically significant at the .05 level, the level of significance for the relationship between faculty

members who would participate in videotaping of classes and who believe the process would be improved by having faculty members examine student ratings was .053, near the accepted .05 value.

Statistical tests notwithstanding, the entire story about faculty attitudes toward the formative evaluation enhancers was not told by examining only the relationships among methods of evaluation and the enhancers. Substantial majorities of respondents agreed that formative evaluation would be improved by training faculty members in methods of evaluation (87.7%), by the establishment of acceptable standards of teaching by the faculty (68.4%), by peer review of student ratings of courses and instructors (62.1%), and by involving the faculty in the pre-planning of formative evaluation (89.6%).

Writers in the field have expressed the view that there are ways that formative evaluation can be enhanced which will increase the likelihood that faculty members will participate (Cohen & McKeachie, 1981; Freer & Dawson, 1985; Heller, 1989; Skoog, 1980). While the results of this study did not confirm a relationship between faculty members' willingness to participate and their agreement with the enhancers, large majorities of faculty members agreed that formative evaluation would be improved by taking these factors into consideration.

Hypothesis 4. Hypothesis 4 was stated as follows: There will be a positive relationship between attitudes toward each method of evaluation and each of the individual/institutional improvement

factors. As shown in Table 16, most relationships between faculty members who would participate in each method of evaluation and their belief that the quality of instruction and student learning would be improved by peer involvement in the process of formative evaluation of instruction were statistically significant. Only the relationships between direct classroom observation and improvement in student learning and between evaluation of instructor-graded student assignments and improvement in the quality of instruction were not significant. Even these levels of significance, .082 and .052, respectively, were near the accepted .05 level.

Writers in the field have indicated that the quality of instruction and student learning would be improved by involving peers in the formative evaluation of instruction of colleagues (Bulcock, 1984; Cancelli, 1987; Hart, 1987). A majority of respondents (72.3%) in Britt's study expressed the view that colleague evaluation could help in improving teaching and student learning. In the present study, large majorities of respondents agreed with the conclusions reached by scholars and the participants in Britt's study.

Except for significant relationships between videotaping and each of the remaining individual/institutional improvement factors (the tenure success of the junior faculty, the morale of the senior faculty, and the collegial climate of the institution), no significant relationships among methods and improvement factors were found.

Table 16

Summary of the Results of the Components of Hypothesis 4

	Method of Formative Evaluation			
Improvement Factor	Direct Classroom Observation	Videotaping of Classes	Evaluation of Course Materials	Evaluation of Graded Assignments
Quality of Instruction	Supported	Supported	Supported	Not Supported
Student Learning	Not Supported	Supported	Supported	Supported
Tenure Success of the Junior Faculty	Not Supported	Supported	Not Supported	Not Supported
Morale of the Senior Faculty	Not Supported	Supported	Not Supported	Not Supported
Collegial Climate of the Institution	Not Supported	Supported	Not Supported	Not Supported

Nonetheless, the entire story about the individual/institutional improvement factors was not told by examining only the relationships between variables. Large proportions of faculty members believed that formative evaluation would result in improvements in the quality of instruction (76.4%), in student learning (57.5%), and in the tenure success of the junior faculty (60.4%). Much smaller proportions of faculty members expressed the belief that such a program would improve the morale of the senior faculty (23.5%) or the collegial climate of the college or university (42.8%).

Recommendations for Practice

A number of conclusions may be drawn from the review of related literature and research and from the analysis of the data from this study. Recommendations for the practice of peer review in the process of formative evaluation of instruction in higher education are based on these conclusions as follows:

1. The literature base for this study revealed that improvement in the quality of teaching in higher education is contingent, in part, on the commitment of administrators to that end. The reward structure of colleges and universities should reflect that commitment.

2. This study and the literature indicate that teaching will be improved by having programs of formative evaluation of instruction, distinct from summative evaluation, in place. Faculty members and academic administrators should work collaboratively to find ways and means to implement direct classroom observation,

videotaping of classes, evaluation of course materials, and evaluation of instructor-graded student assignments as methods in the process of formative evaluation of instruction.

3. This study indicates that quality of instruction and student learning might be improved by involving peers in the formative evaluation of the teaching performance of colleagues. Since teaching and learning are "the business of the business," it is in the best interests of students, the faculty, and colleges and universities to have such programs of formative evaluation in place. Besides making a commitment to improve the quality of teaching, administrators should provide moral and financial support and released time to faculty so that they will develop and participate in programs of this type.

4. Teaching performance can be evaluated directly and indirectly. Although some aspects of teaching can be evaluated only by direct classroom observation and/or by studying videotapes of instruction, pre-interactive and post-interactive teaching events can be more effectively assessed by evaluating course materials and instructor-prepared and instructor-graded student assignments. Programs of formative evaluation should include methods designed to evaluate not only events which occur while an instructor and students are interacting, but also those aspects of teaching which occur prior to and following delivery of instruction. The program should include all of the following methods: direct classroom observation, videotaping of classes, evaluation of course materials, and evaluation of instructor-graded student assignments.

5. Faculty members in this study did not believe that (a) their academic freedom will be circumscribed by the methods investigated, (b) these methods may not measure accurately their teaching performance, (c) these methods may not evaluate a typical teaching and learning situation, or (d) these methods may result in subjective, rather than objective, assessments of their performance. Perhaps they should be aware that any method used by itself may not measure accurately their performance and/or may not evaluate a typical teaching and learning situation. That caveat notwithstanding, faculty members who participate in voluntary programs of direct classroom observation, videotaping of classes, evaluation of course materials, evaluation of instructor-graded student assignments, or, especially, a combination of these methods of evaluation, can gain insights into the process of teaching and learning which will help them improve their teaching. Faculty members should be encouraged by their peers and by academic administrators to take part in comprehensive programs of formative evaluation of instruction.

6. Providing training to faculty members in methods of evaluation may help them to become more comfortable in assessing the strengths and weaknesses of their colleagues' teaching. It may also give faculty members insights into the process of teaching and learning that they have not previously considered. Administrators should find ways to provide training of this type.

7. It is neither necessary nor desirable for the faculty to establish acceptable standards of teaching, especially if that implies a single set of standards to be applied in the evaluation of all faculty in all situations. According to the review of literature, effective teaching is contingent on the complex interrelationship of environmental and personal factors. In formative evaluation, the strengths and weaknesses of a professor's performance should be assessed in the broadest possible context, one that includes the physical-temporal setting, the specifics of the situation, the linguistic dimension, the dramatalurgical and sociological settings, the curricular context, and classroom assessment procedures.

8. Faculty members understand the circumstances under which student ratings of courses and instructors are obtained. The process of formative evaluation should include faculty peer review of student ratings in light of these circumstances. Faculty members should familiarize themselves with the literature on student ratings of courses and instructors.

9. Faculty members are more likely to participate in formative evaluation if they are involved in the pre-planning of the program. If they are convinced that the program is theirs, they are more likely to be committed to the ultimate goals of improvements in the quality of instruction and in student learning. While participation should be voluntary, academic administrators should encourage participation by emphasizing the importance of effective teaching

and by making time available to faculty members for the development of and participation in formative evaluation.

10. It is nearly impossible to separate the concepts of morale and collegiality when they are discussed in connection with formative evaluation. While scholars insist that faculty morale and the collegial climate of the college or university will improve by faculty participation in effective programs of formative evaluation, faculty members who took part in this study remained doubtful. Faculty members who are convinced that morale and collegiality can be improved by faculty participation in programs of this type will have to "sell" the idea to their skeptical colleagues.

11. Participants in this study believed that the tenure success of the junior faculty will improve by having non-tenured faculty participate in formative evaluation of instruction. Senior faculty members and administrators should encourage non-tenured faculty members to take part in these programs. Mentorship programs involving tenured faculty members who are recognized as effective teachers and non-tenured faculty members should be considered as one way to implement such a program.

Recommendations for Further Research

A number of further studies have been suggested from the research conducted in connection with this study. Studies like the following are recommended:

1. This study might be replicated with samples of the population from other types of colleges and universities (e.g.,

institutions affiliated with the American Association of State Colleges and Universities or The Renaissance Group, junior and community colleges, and institutions in other areas across the country).

2. This study might be replicated using a larger sample or the population from a consortium of colleges and universities so that demographic breakdowns of data can be more adequately investigated.

3. Ethnographic studies might be conducted so that the problem investigated in this study can be approached more intensively.

4. Experimental studies might be conducted at colleges and universities where programs of formative evaluation are in place. Since programs of this type are available at two of the institutions affiliated with the Iowa Association of Independent Colleges and Universities, it might be possible for researchers to conduct studies of this type at those institutions.

5. Longitudinal studies on the effects of formative evaluation involving colleagues as evaluators might be conducted. To date, only the short-term effects of formative evaluation have been studied.

6. Experimental studies might be conducted at colleges and universities where formative evaluation is required of faculty members. To date, only voluntary programs have been studied.

7. Studies in which respondents list in order of priority the methods in which they would participate might be conducted.

Knowing which methods faculty members consider the most effective could be used to gain initial support for what eventually could become a more comprehensive program of formative evaluation.

8. Detractors other than the four considered in this study should be investigated. Because detractors investigated in this study, apparently do not deter faculty members from participating in formative evaluation, researchers should continue to look for other reasons why formative evaluation is not normally a part of comprehensive faculty evaluation programs.

9. Enhancers other than the four considered in this study should be investigated. Researchers should continue to search for ways to make formative evaluation attractive enough to faculty members so that the faculty will develop and participate in such programs.

10. Studies should be conducted in which the relationship between participation in various methods of formative evaluation and the tenure success of junior faculty members is studied. Surprisingly, no reference to this relationship was found in the literature.

11. Researchers should construct experimental studies of programs of formative evaluation in which the differential effects of evaluators from the same (or similar) and unrelated disciplines are investigated.

12. A means to get more forthright responses to self-report instruments like the questionnaire used to obtain data for this

study should be investigated. While respondents probably did not intentionally provide inaccurate answers to questionnaire items, one wonders why there was such widespread support for the methods of formative evaluation investigated and so few programs of this type in place at these colleges and universities. Perhaps participants have provided professionally-acceptable, rather than completely honest, responses.

Conclusion

A review of the literature shows that many scholars have recommended that attention be given to the evaluation of college teaching. Many of these scholars have recommended that formative evaluation--assessment specifically designed to improve instruction--be put into place to complement evaluation conducted for the purpose of decision-making regarding reappointment, promotion, tenure, and compensation.

Large majorities of faculty members participating in this study believe that direct classroom observation, videotaping of classes, evaluation of course materials, and evaluation of instructor-graded student assignments, when conducted by peers for the purpose of formative evaluation, would help them improve their teaching. Given that both scholars and the faculty members represented in this study have agreed that formative evaluation would result in improvement in the quality of instruction and may have other benefits, this researcher is convinced that the time has come for colleges and universities to implement--and then systematically evaluate--programs of formative evaluation.

References

Aleamoni, L. M. (1981). Standards for evaluation of instruction (Note to the Faculty, Number 11). Tucson: University of Arizona Institutional Research and Development.

Aleamoni, L. M. (1984). Peer evaluation (Note to the faculty, Number 15). Tucson: University of Arizona Institutional Research and Development.

Aleamoni, L. M., & Yimer, M. (1973). An investigation of the relationship between colleague rating, student rating, research productivity, and academic rank in rating instructional effectiveness. Journal of Educational Psychology, 64(3), 274-277.

Arden, E. (1989, Summer). Who should judge the faculty? The College Board Review, 37-39.

Astin, A. W., & Lee, C. B. T. (1966). Current practices in the evaluation and training of college teachers. The Educational Record, 47, 361-375.

Ballard, M. J., Reardon, J., & Nelson, L. (1976). Student and peer ratings of faculty. Teaching of Psychology, 3, 88-91.

Batista, E. E. (1976). The place of colleague evaluation in the appraisal of college teaching: A review of the literature. Research in Higher Education, 4, 257-271.

Bell, M. E., Dobson, E. C., & Gram, J. M. (1977). Peer evaluation as a method of faculty development. Journal of the College and University Personnel Administration, 28(4), 15-17.

Bergman, J. (1979). The effectiveness of peer ratings at the university level. Journal of Teaching and Learning, 4(3), 34-37.

Bergman, J. (1980). Peer evaluation of university faculty. College Student Journal (monograph ed.), 14(3, Pt. II), 1-21.

Bergquist, W. H., & Phillips, P. (1975). A handbook for faculty development. New York: Danville.

Bibliography on peer evaluation. (1985). Instructional Evaluation, 8(1), 22-24.

Blackburn, R. T., & Clark, M. T. (1975). An assessment of faculty performance: Some correlates between administrator, colleague, student and self-ratings. Sociology of Education, 48, 342-356.

Brandenburg, D. C., Braskamp, L. A., & Ory, J. C. (1979). Considerations for an evaluation program of instructional quality. CEDR Quarterly, 12(4), 8-12.

Braskamp, L. A. (1978). Colleague evaluation of instruction. Faculty Development and Evaluation in Higher Education, 4, 1-9.

Britt, N., Jr. (1982). Faculty attitudes about colleague evaluation of teaching. Dissertation Abstracts International, 42, 5034A. (University Microfilms No. 82-09886)

Brock, S. C. (1981). Evaluation-based teacher development. In J. Millman (Ed.), Handbook of teacher evaluation (pp. 224-243). Beverly Hills, CA: Sage.

Bryant, P. T. (1967). By their fruits ye shall know them. Journal of Higher Education, 38, 326-330.

Bulcock, J. W. (1984). Why can't we define good teaching? Paper presented at the Annual Meeting of the Canadian Society for the Study of Education, Guelph, Ontario. (ERIC Document Reproduction Service No. ED 248 207)

Cancelli, A. (1987). Methods for arriving at clinical judgments in peer evaluation. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC. (ERIC Document Reproduction Service No. ED 282 914)

Carroll, J. G., & Goldberg, S. R. (1989). Teaching consultants: A collegial approach to better teaching. College Teaching, 37, 143-146.

Ceci, S. J., & Peters, D. P. (1982). Peer review: A study of reliability. Change, 14(6), 44-48.

Centra, J. A. (1975). Colleagues as raters of classroom instruction. Journal of Higher Education, 46, 327-337.

Centra, J. A. (Ed.). (1977). Reviewing and evaluating teaching. San Francisco: Jossey-Bass.

Centra, J. A. (1986). Colleague evaluation: The critical link. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. 275 722)

Cohen, J. (1977). Statistical power analysis for the behavioral sciences (rev. ed.). New York: Academic Press.

Cohen, P. A., & McKeachie, W. J. (1981). The role of colleagues in the evaluation of college teaching. Improving College and University Teaching, 28, 147-154.

Cowen, D. L., Davis, G. L., & Bird, S. E. (1976). Peer review in medical education. Journal of Medical Education, 51, 130-131.

Craig, J. R., Redfield, D. L., & Galluzzo, G. R. (1986). Evaluating effective teaching in colleges and universities: How far have we come? Paper presented at the Annual Meeting of the American Evaluation Association, Kansas City, MO. (ERIC Document Reproduction Service No. 282 888)

Cross, K. P. (1986). Using assessment to improve instruction. Cambridge, MA: Harvard University. (ERIC Document Reproduction Service No. ED 284 896)

Dienst, E. R. (1981). Evaluation by colleagues. San Francisco: University of California. (ERIC Document Reproduction Service No. 209 341)

Doyle, K. O., Jr., & Crichton, L. I. (1978). Student, peer, and self evaluations of college instruction. Journal of Educational Psychology, 70, 815-826.

Dressel, P. L. (1976). Faculty. In P. L. Dressel, Handbook of academic evaluation (pp. 331-375). San Francisco: Jossey-Bass.

Eckert, R. E. (1950). Ways of evaluating college teaching. School and Society, 71, 65-69.

Edwards, S. (1974). A modest proposal for the evaluation of teaching. Liberal Education, 60, 316-326.

Elbow, P. (1980). One-to-one faculty development. New Directions for Teaching and Learning, 4, 25-40.

Erickson, G. G., & Erickson, B. L. (1979). Improving college teaching. Journal of Higher Education, 50, 670-683.

Farmer, C. H. (1976). Colleague evaluation: The silence is deafening. Liberal Education, 62, 432-436.

Fitzgerald, M. J., & Grafton, C. L. (1981). Comparisons and implications of peer and student evaluations for a community college faculty. Community/Junior College Research Quarterly, 5, 331-337.

Freer, M., & Dawson, J. (1985). DON'T evaluate your teachers. Phi Delta Kappan, 66, 720-722.

French-Lazovik, G. (1981). Peer review: Documentary evidence in the evaluation of teaching. In J. Millman (Ed.), Handbook of teacher evaluation (pp. 73-89). Beverly Hills, CA: Sage.

Gage, N. L. (1961). The appraisal of college teaching. Journal of Higher Education, 32, 17-22.

Galm, J. A. (1985). Welcome to post-tenure review. College Teaching, 33, 65-67.

Goldhammer, R. (1969). Classical supervision. New York: Holt, Rinehard and Winston.

Greenwood, G. E., & Ramagli, H. J., Jr. (1980). Alternatives to student ratings of college teaching. Journal of Higher Education, 51, 673-684.

Gunn, B. (1982). Evaluating faculty performance: A holistic approach. Journal of the College and University Personnel Association, 34(4), 23-30.

Guthrie, E. R. (1949). The evaluation of teaching. The Educational Record, 30, 109-115.

Hart, F. R. (1987). Teachers observing teachers. In J. H. Broderick (Ed.), Teaching at an Urban University. Boston: University of Massachusetts at Boston (pp. 15-24). (ERIC Document Reproduction Service No. ED 290 704)

Heller, D. A. (1989). Peer supervision: A way of professionalizing teaching. Bloomington, IN: Phi Delta Kappa.

Hodgkinson, H. (1972). Unlock the doors: Let your colleagues in: Faculty reward and assessment systems. In J. Braun & T. A. Emmet (Eds.), The academic department and divison chairman (pp. 208-215). Detroit: Balamp.

Hoyt, D. P., & Howard, G. S. (1978). The evlauation of faculty development programs. Research in Higher Education, 8, 25-38.

Jones, M. A. (1986). Participatory evaluation of a departmental peer review process for awarding merit pay to university faculty. Dissertation Abstracts International, 48, 0316A. (University Microfilms No. 87-11724)

Keig, L. W. (1989). Faculty evaluation: Iowa Association of Independent Colleges and Universities. Unpublished manuscript.

Lee, B. A. (1982). Balancing confidentiality and disclosure in faculty peer review: Impact on Title VII litigation. Journal of College and University Law, 9, 279-314.

Levinson-Rose, J., & Menges, R. J. (1981). Improving college teaching: A critical review of research. Review of Educational Research. 3, 403-434.

Lichty, R. W., & Peterson, J. M. (1979). Peer evaluations--A necessary part of evaluating teaching effectiveness. Duluth: University of Minnesota. (ERIC Document Reproduction Service No. ED 175 352)

Maslow, A. H., & Zimmerman, W. (1956). College teaching ability, scholarly activity and personality. The Journal of Educational Psychology, 47, 185-189.

Mathias, H., & Rutherford, D. (1982a). Course evaluation at Birmingham: Some implications for the evaluation of university teaching. Studies in Educational Evaluation, 7, 263-266.

Mathias, H., & Rutherford, D. (1982b). Lecturers as evaluators: The Birmingham experience. Studies in Higher Education. 7(1), 47-56.

McCarthey, S. J., & Peterson, K. D. (1988). Peer review of materials in public school teacher evaluation. Journal of Personnel Evaluation in Education, 1, 259-267.

McIntosh, T. H., & Van Koevering, T. E. (1986). Six-year case study of faculty peer review, merit ratings, and pay awards in a multidisciplinary department. Journal of the College and University Personnel Association, 37, 5-14.

McIntyre, C. J. (1978). Peer evaluation of teaching. Urbana-Champaign: University of Illinois. (ERIC Document Reproduction Service No. ED 180 295)

McIntyre, K. E. (1986). Using classroom observation data for diagnosis purposes. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED 275 731)

McKeachie, W. J. (1986). Teaching tips: A guidebook for beginning college teachers. Lexington, MA: Heath.

McKeachie, W. J. (1987). Can evaluating instruction improve teaching? In L. M. Aleamoni, Techniques for evaluating and improving instruction (pp. 3-7). San Francisco: Jossey-Bass.

Menges, R. J. (1985). Career-span faculty development. College Teaching, 33, 181-184.

Mikula, A. R. (1979). Using peers in instructional development. Altoona: The Pennsylvania State University. (ERIC Document Reproduction Service No. 172 599)

Miller, L. H., Jr. (1990). Hubris in the academy. Change, 22(5), 9-11, 53.

Millman, J. (Ed.). (1981). Handbook of teacher evaluation. Beverly Hills, CA: Sage.

Murray, H. G. (1975). Predicting student ratings of college teaching from peer ratings of personality types. Teaching of Psychology, 2(2), 66-69.

Parsons, T. (1954). The social system. New York: The Free Press.

Pew Higher Education Research Program. (1989). The business of the business. Policy Perspectives, 1(3), 1-7.

Pew Higher Education Research Program. (1990). Back to business. Policy Perspectives, 3(1), 1-8.

Prater, D. L. (1983). What counts as effective university teaching: The state of the art. Paper presented at the Annual Meeting of the Southwest Educational Research Association, Houston, TX. (ERIC Document Reproduction Service No. ED 227 149)

Razor, J. E. (1979). The evaluation of administrators and faculty members--Or evaluation of the "boss" or each other. Normal: Illinois State University. (ERIC Document Reproduction Service No. 180 355)

Riegle, R. P., & Rhodes, D. M. (1986). Avoiding mixed metaphors in faculty evaluation. College Teaching, 34, 123-128.

Romberg, E. (1985). Description of peer evaluation within a comprehensive evaluation program in a dental school. Instructional Evaluation, 8(1), 10-16.

Root, L. S. (1987). Faculty evaluation: Reliability of peer assessment of research, teaching, and service. Research in Higher Education, 26, 71-84.

Roper, S. S., Deal, T. E., & Dornbusch, S. (1976). Collegial evaluation of classroom teaching: Does it work? Educational Research Quarterly, 1(1), 56-66.

Sauter, R. C., & Walker, J. K. (1976). A theoretical model for faculty "peer" evaluation. American Journal of Pharmaceutical Education, 40, 165-166.

Schneider, L. S. (1975). Faculty opinion of the spring 1974 peer evaluation. Los Angeles: Los Angeles City College. (ERIC Document Reproduction Service No. 104 493)

Scriven, M. S. (1980). The evaluation of college teaching. Syracuse, NY: National Council on States Inservice Education. (ERIC Document Reproduction Service No. 203 729)

Scriven, M. S. (1983). Evaluation ideologies. In G. F. Madaus, M. S. Scriven, & D. L. Stufflebeam (Eds.), Evaluation models: Viewpoints on educational and human services evaluation (pp. 242-249). Boston: Kluver-Nijhof.

Scriven, M. S. (1985). New frontiers of evaluation. Evaluation Practices, 7(1), 7-44.

Seldin, P. (1984). Changing practices in faculty evaluation. San Francisco: Jossey-Bass.

Shatzky, J., & Silberman, R. (1986). Master-students: A teaching technique. Journal of College Science Teaching, 16, 119-120.

Singh, R. (1984). Peer-evaluation: A process that could enhance the self-esteem and professional growth of teachers. Education, 105(1), 73-75.

Skoog, G. (1980). Improving college teaching through peer observation. Journal of Teacher Education, 31(2), 23-25.

Smith, A. (1985). The challenge of peer evaluation. Instructional Evaluation, 8(1), 2-3.

Smith, G. (1987). The practitioners of staff development. Journal for Higher Education, 11, 58-67.

Smith, M. R. (1981). Protecting the confidentiality of faculty peer review records: Department of Labor v. The University of California. Journal of College and and University Law, 8, 20-53.

Smith, P., Hausken, C., Kovacevich, H., & McGuire, M. (1988). Alternatives for developing teacher effectiveness. Seattle, WA: School of Education, Seattle Pacific University. (ERIC Document Reproduction Service No. ED 301 115)

Snedecor, G. W., & Cochran, W. G. (1967). Statistical methods (sixth ed.). Ames, IA: Iowa State University Press.

Soderberg, L. O. (1985). Dominance of research and publications: An unrelenting tyrrany. College Teaching, 33, 168-172.

Soderberg, L. O. (1986). A credible model: Evaluating classroom teaching in higher education. Instructional Evaluation, 8(2), 13-27.

Sorcinelli, M. D. (1984). An approach to colleague evaluation of classroom instruction. Journal of Instructional Development, 7(4), 11-17.

Spaights, E., & Bridges, E. (1986). Peer evaluations for salary increases and promotions among college and university faculty members. North Central Association Quarterly, 60, 403-410.

Stevens, J. J. (1985). Legal issues in the use of peer evaluation. Instructional Evaluation, 8(1), 17-21.

Stevens, J. J., & Aleamoni, L. M. (1985). Issues in the development of peer evaluation systems. Instructional Evaluation, 8(1), 4-9.

Stodolsky, S. S. (1984). Teacher evaluation: The limits of looking. Educational Researcher, 13(9), 11-18.

Stumpf, W. E. (1980). Peer review. Science, 207, 822-823.

Swanson, R. A., & Sisson, D. J. (1971). The development, evaluation, and utilization of departmental faculty appraisal system. Journal of Industrial Teacher Education, 9(1), 64-79.

Sweeney, J. M. W. (1976). A report on the development and use of a faculty peer evaluation/development program. Dissertation Abstracts International, 37, 5458A. (University Microfilms No. 76-30, 408)

Sweeney, J. M., & Grasha, A. F. (1979). Improving teaching through faculty development triads. Educational Technology, 19, 54-57.

Ururoglu, M. E., & Dwyer, M. M. (1981). Staff review system. Improving College and University Teaching, 29, 121-124.

Ward, M. D., Clark, D. C., & Harrison, G. V. (1981). The observation effect in classroom visitation. Macomb: Western Illinois University. (ERIC Document Reproduction Service No. 204 384)

Webb, W. B. (1955). The problem of obtaining negative nominations in peer ratings. Personnel Psychology, 8, 61-63.

Weinbach, R. W., & Randolph, J. L. (1984). Peer review for tenure and promotion in professional schools. Improving College and University Teaching, 32, 81-86.

Wherry, R. J., & Fryer, D. H. (1945). Buddy ratings: Popularity contest or leaders in criteria. Personnel Psychology, 2, 147-159.

Wilson, R. C., Dienst, E. R., & Watson, N. L. (1973). Characteristics of effective college teachers as perceived by their colleagues. Journal of Educational Measurement, 10, 31-37.

Wood, P. H. (1977). The description and evaluation of a college department's faculty rating system. Bowling Green, OH: Bowling Green State University. (ERIC Document Reproduction Service No. 142 128)

Wood, P. H. (1978). Student and peer ratings of college teaching and peer ratings of research and service: Four years of departmental evaluation. Bowling Green, OH: Bowling Green State University. (ERIC Document Reproduction Service No. 155 218)

Zirkel, P. A. (1985). Faculty review in the promotion and tenure process beyond the departmental level. Planning for Higher Education, 13(4), 15-17.

APPENDICES

APPENDIX A

SAMPLE SIZE FORMULA

Sample Size Formula

Snedecor and Cochran's (1967) formulas for sample size and the finite population correction were used.

Formula

$$n = \frac{(1.96^2 * P * (100 - P))}{r^2}$$

Where:

1. P is the expected percentage of the elements in the universe having the characteristic under study (worst case of 50% is assumed).
2. r is the reliability of the sample (e.g., $\pm$ r%)
3. n is the sample size needed.

Correction

$$n = \frac{(U * n)}{(U + n)}$$

Where:

1. U is the size of the universe.

APPENDIX B

IAICU Institutions Offering the Baccalaureate Degree

IAICU Institutions Offering the Baccalaureate Degree

COLLEGE OR UNIVERSITY	FULL-TIME FACULTY	PROPORTION OF POPULATION	NUMBER IN SAMPLE
Briar Cliff College	68	.033	25
Buena Vista College	70	.034	25
Central College	95	.046	34
Clarke College	62	.030	22
Coe College	78	.038	28
Cornell College	74	.036	27
Dordt College	79	.038	28
Drake University	227	.110	82
University of Dubuque	58	.028	21
Graceland College	57	.027	20
Grand View College	66	.032	24
Grinnell College	163	.079	59
Iowa Wesleyan College	42	.020	15
Loras College	124	.060	45
Luther College	166	.080	60
Marycrest College	55	.027	20
Morningside College	78	.038	28
Mount Mercy College	36	.017	13
Mount Saint Clare College	64	.031	23
Northwestern College	52	.025	19

(continued)

COLLEGE OR UNIVERSITY	FULL-TIME FACULTY	PROPORTION IN POPULATION	NUMBER IN SAMPLE
Saint Ambrose College	102	.049	37
Simpson College	73	.035	29
Upper Iowa University	26	.013	10
Wartburg College	81	.039	29
Westmar College	38	.018	14
William Penn College	39	.019	14

APPENDIX C

Letter of Transmittal

University of Northern Iowa
Center for Social and Behavioral Research
Cedar Falls, Iowa 50614-0286
Telephone (319) 273-2105

October 9, 1990

Dear Faculty Member:

We are conducting a study among a sample of faculty who teach at institutions that are affiliated with the Iowa Association of Independent Colleges and Universities. The purpose of our study is to assess faculty opinion on matters relating to the evaluation of an instructor's teaching performance. The procedures involved in such an assessment are described in the enclosed questionnaire. Your name was selected at random for inclusion in our study.

We would appreciate your completing and returning the questionnaire in the postage paid envelope. How you personally respond to the questions will only be known to you since we do not ask for your name.

By your participating in this study you will be providing valuable information that can be used by educators, administrators, and others in making decisions regarding peer involvement in the assessment of instruction. In addition, the data will be used in a doctoral dissertation study. We value your thoughts and opinions on this important matter, and appreciate your participating in this study.

Sincerely,

Larry Keig

Larry W. Keig
Project Director

Enclosures (2)

APPENDIX D

Questionnaire

CENTER FOR SOCIAL AND BEHAVIORAL RESEARCH
University of Northern Iowa

Survey of Attitudes Regarding Evaluation of Instruction

Directions: This questionnaire seeks your opinions on issues relating to "formative evaluation of instruction" by "informed peers." That is, colleagues who are knowledgeable (by training and/or experience) in the areas of teaching strategies, learning strategies, methods of instruction, and instructional materials among others, evaluate a faculty members methods of instruction for purposes of improving that person's instructional techniques.

This type of evaluation is used *only* for improving instruction. It is *not* used in decisions concerning reappointment, promotion, tenure, and/or compensation. Findings from this type of evaluation are normally shared through discussions between/among evaluators and the faculty member being evaluated, and are not usually communicated solely in writing or numerically from rating scales.

Please use the above description of "formative evaluation" when answering the following questions.

Which of the following types of instructional evaluations, if any, would you agree to participate in if the evaluation was performed by an informed peer?

Yes	No	Not Sure	Types of Evaluations
☐	☐	☐	Direct classroom observation.
☐	☐	☐	Videotaping of classes.
☐	☐	☐	Evaluation of course materials (e.g., syllabi, grading practices, reading lists, handouts).
☐	☐	☐	Evaluation of instructor-graded student assignments (e.g., tests, term papers, projects, homework).

The following statements relate to the four types of instructional evaluations listed on the preceding page. Please read each statement in relation to the instructional evaluation category listed, and circle whether you *Strongly Agree*, (A), *Agree*, are *Uncertain* (U), *Disagree* (D),or *Strongly Disagree* (SD) with the statement.

Direct classroom observation would . .

infringe on my academic freedom.	SA	A	U	D	SD
probably not measure accurately my teaching performance.	SA	A	U	D	SD
probably not be representative of a typical teaching-learning situation.	SA	A	U	D	SD
result in a subjective, rather than objective, assessment of my performance.	SA	A	U	D	SD

Videotaping my classes would . . .

infringe on my academic freedom.	SA	A	U	D	SD
probably not measure accurately my teaching performance.	SA	A	U	D	SD
probably not be representative of a typical teaching-learning situation.	SA	A	U	D	SD
result in a subjective, rather than objective, assessment of my performance.	SA	A	U	D	SD

Evaluating my course materials would . . .

infringe on my academic freedom.	SA	A	U	D	SD
probably not measure accurately my teaching performance.	SA	A	U	D	SD
probably not be representative of a typical teaching-learning situation.	SA	A	U	D	SD
result in a subjective, rather than objective, assessment of my performance.	SA	A	U	D	SD

Evaluating my graded student assignments would . . .

infringe on my academic freedom.	SA	A	U	D	SD
probably not measure accurately my teaching performance.	SA	A	U	D	SD
probably not be representative of a typical teaching-learning situation.	SA	A	U	D	SD
result in a subjective, rather than objective, assessment of my performance.	SA	A	U	D	SD

The following statements relate to items which enhance the process of instructional evaluation. Please read each statement and circle whether you *Strongly Agree*, (A), *Agree*, are *Uncertain* (U), *Disagree* (D),or *Strongly Disagree* (SD) with the statement.

A program of instructional evaluation in which informed peers evaluate the teaching performance of colleagues would be increased if . . .

training in evaluation methods was made available to the faculty.	SA	A	U	D	SD
acceptable standards of teaching were established by the faculty.	SA	A	U	D	SD
student ratings of participants were included.	SA	A	U	D	SD
parties involved in the process were consulted in the pre-planning.	SA	A	U	D	SD

A program of instructional evaluation in which informed peers evaluate the teaching performance of colleagues would improve . . .

the quality of instruction.	SA	A	U	D	SD
student learning.	SA	A	U	D	SD
the tenure success of junior faculty members.	SA	A	U	D	SD
the morale of senior faculty members.	SA	A	U	D	SD
the collegial climate of the college/university.	SA	A	U	D	SD

(OVER, PLEASE)

In order to have a better understanding of faculty members participating in our study, we would appreciate your answering the following background information questions.

What is your sex? ☐ Male ☐ Female

What was your age on your last birthday? _____

For approximately how many years have your been teaching at the college or university level? (If less than one year, please enter a zero (0).) _____ Years

What is your current academic rank?

☐ Professor
☐ Associate Professor
☐ Assistant Professor
☐ Instructor (or equivalent)
☐ Adjunct/part-time instructor
☐ Other (Please specify:)

_____

Are you tenured or non-tenured? ☐ Tenured ☐ Non-tenured

Do you teach on a semester or quarterly hour basis?

☐ Semester ☐ Quarterly ☐ Other (Specify:) _____

How many credit hours do you teach during a typical semester or quarter? _____ Hours

Other (Specify:) _____

In what academic discipline do you teach all or most of your courses?

Discipline: _____

-- THANK YOU FOR PARTICIPATING IN THIS SURVEY --

APPENDIX E

Academic Disciplines by Categories

Academic Disciplines by Categories

Arts/Humanities

00 Humanities
01 English
02 Philosophy
03 Communications and Speech
04 Theology and Religion
05 Foreign Language
06 Music
07 German
08 Journalism
09 Art
10 Classics
11 Linguistics
12 Theatre
13 Spanish
14 French
15 Composition

Health Care

20 Nursing
21 Pharmacology

Math and Science

30 Information Services
31 Computer Science
32 Mathematics
33 Physics
34 Chemistry
35 Science
36 Biology
37 Engineering
38 Agriculture
39 Natural Science
40 Industrial Technology
41 Statistics
42 Astronomy
43 Geology
44 Home Economics

Social Sciences

50 Sociology
51 Social Work
52 Psychology
53 Political Science
54 History
55 Social Science
56 Anthropology

Health, P.E., and Recreation

60 Physical Education
61 Health, Training
62 Recreation

Education

70 Education
71 Special Education
72 Early Childhood Education
73 Elementary Education
74 Library Science

Business

80 Office Administration
81 Accounting
82 Business Administration
83 Economics
84 Management
85 Marketing
86 Finance
87 Consumer Education

Miscellaneous

98 Not indicated
99 Other